Incidental learning of faces during threat: No evidence for increased autonomic arousal to “unrecognized” threat identities

Sabine Schellhaas (sabine.schellhaas@zi-mannheim.de)
Central Institute of Mental Health
Christian Schmahl
Central Institute of Mental Health
Florian Bublatzky
Central Institute of Mental Health

Article

Keywords: threat of shock, recognition, source memory, MPT, startle reflex, skin conductance response

Posted Date: May 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1619038/v1

Abstract

Remembering an unfamiliar person and the contextual conditions of that encounter is important for adaptive future behavior, especially in a potentially dangerous situation. Initiating defensive behavior in the presence of former dangerous circumstances can be crucial. Recent studies showed selective electrocortical processing of faces that were previously seen in a threat context compared to a safety context, however, this was not reflected in conscious recognition performance. Here, we investigated whether previously seen threat-faces, that could not be remembered, were capable to activate defensive psychophysiological response systems. During an encoding phase, 50 participants with low to moderate levels of anxiety (partially inpatients) viewed 40 face pictures with neutral expressions (6s each), without an explicit learning instruction (incidental learning task). Each half of the faces were presented with contextual background colors that signaled either threat-of-shock or safety. In the recognition phase, all old and additional new faces (total of 60) were presented intermixed without context information. Participants had to decide whether a face was new or had previously been presented in a threatening or a safe context. Results show moderate face recognition independent of context conditions. Startle reflex and skin conductance response (SCR) were more pronounced for threat compared to safety during encoding. For SCR, this differentiation was enhanced with higher levels of depression and anxiety. There were no differential startle reflex or SCR effects during recognition. From a clinical perspective, these findings do not support the notion that perceptual biases and physiological arousal directly relate to threat-associated identity recognition deficits in healthy and clinical participants with anxiety and trauma-related disorders.

Introduction

Humans have difficulties in identifying unfamiliar persons, especially when the situational context changed [1]–[3]. This can become dangerous, e.g., when a person encountered in a hostile situation is met again but not recognized. During initial encounter, defensive response programs are activated when an evolutionary prepared aversive situation occurs (e.g., fear, freezing, fight/flight; [4]). In the second encounter, when the unfamiliar person is not identified, the aversive context of the first encounter is hardly recognized at all [5]. However, the question remains whether the unrecognized person is nevertheless associated with the threatening context even though neither is remembered. Thus, the present research question is whether we find defense activation (i.e., enhanced skin conductance response [SCR] and startle reflex) to non-recognized persons who were previously encountered in a threatening context.

Psychophysiological defense mechanisms have been suggested to operate relatively independently from conscious cognitive processing of stimulus associations [4]. For instance, on the neural level we observed selective processing of previously seen faces as a function of whether faces were encoded in a threatening or safe context [6]. Interestingly, this differential old/new-ERP effect was not reflected in conscious recognition performance and resulted in poor face and context memory. Such difficulties in recognizing unfamiliar faces seen only briefly (e.g., for 1s or 6s) have also been observed in other studies in which a large number of ninety faces were presented in a blocked manner [5]. Participants did not know whether and, possibly more importantly, under what conditions they had seen a person before (i.e., during threat or safe conditions). Because neural processing is assumed to have direct access to defensive response programs irrespective of conscious face and context recognition (e.g., [7]) the question emerged whether threat-selective perceptual processing results in priming of defensive psychophysiological response patterns to unrecognized threat relative to safety associated faces.

From a clinical perspective this is an important question as perceptual biases and physiological arousal may contribute to threat-associated identity recognition deficits in (socially) anxious and/or traumatized participants [8]–[9]. In this context, recent research suggested attentional threat biases and hypervigilance for highly anxious individuals [10]–[11]. Moreover, the generalization of the defensive startle reflex activity is a central feature of pathological anxiety especially for posttraumatic stress disorders [12]–[14]. However, it remains unclear whether threat-related overgeneralization and attentional biases extend to the memory processes and remain constant once the acute threat has disappeared.

The present study examined whether participants show threat-enhanced activity of the autonomic and somatic nervous system (i.e., enhanced SCRs and threat-potentiated startle reflex) to people that were previously met in a threat context but cannot be remembered. To this end, an item/source (i.e., face/context) recognition task was combined with the threat-of-shock paradigm, in which participants were verbally informed about the possibility to receive electrical shocks when a particular colored background was present (e.g., blue signals threat) whereas another background color indicated safety (e.g., green signals safety). During an initial encoding phase (incidental learning), unfamiliar faces were paired with either a threat or safety context. It is assumed that defensive responses are increased for faces in a threat relative to a safe context (i.e., threat-potentiated startle and SCR; [15]–[16]). Moreover, startle potentiation should be more pronounced in more (socially) anxious participants while the difference between threat and safety decreases (generalization, [12], [17]).

During the following recognition phase, an unexpected face/context recognition task was performed [6], [5]. Participants indicated whether they had seen a face with a threat or safety background, or whether it was new. Despite the notion of a memory enhancing effect of aversive apprehension ([18]–[19]), we expected rather poor recognition performance for incidentally learned faces [6], [5], [2]. Of particular interest is the physiological responding to faces that had been seen previously within threat compared to safety background. According to the notion of a threat advantage [4], faces encoded during threat should elicit enhanced SCRs and startle reflex potentiation during recognition relative to safely encoded faces. Similar to encoding, more pronounced SCR and startle reflex are expected for (socially) anxious [20] and traumatized participants with more severe anxiety [17]. For these participants, less face differentiation (old-threat vs. old-safe vs. new) was expected, reflecting maladaptive evaluative processing of threatening face–context compounds.

Methods

Participants

Fifty participants (35 female, 12 male, 3 other) between the age of 18 to 51 years (M = 24.87 [SD = 7.46]) were recruited from the University of Mannheim, the SRH Heidelberg and the general population of Mannheim (Germany). For recruitment of a diverse sample, a transdiagnostic and dimensional approach was followed. To increase the number of high-anxious participants with clinically relevant psychopathology, thirteen participants were recruited from the Department of Psychosomatic Medicine and Psychotherapy at the Central Institute of Mental Health in Mannheim. Of all participants, 45.8% self-reported a current or past mental disorder. 31.3% of these participants self-reported current or past Borderline Personality Disorder (BPD), 29.2% posttraumatic stress disorder (PTSD), 25% depression and 8.3% anxiety disorders. 14.3% self-reported another mental disorder (including attention deficit/hyperactivity disorder, (atypical) anorexia nervosa, obsessive-compulsive disorder, narcissistic personality disorder).

We assessed anxiety, depression and trauma measures (STAI-S = 38.88 [SD = 13.23], STAI-T = 49.68 [SD = 13.93], SPIN = 40.41 [SD = 15.68], BDI II = 41.03 [SD = 16.57], CTQ = 51.71 [SD = 22.72], using the German versions of the Social Phobia Inventory, SPIN, [21]; State-Anxiety Inventory, STAI-T/S, [22], Beck Depression Inventory II, BDI II, [23], Childhood Trauma Questionnaire, CTQ, [24]). Participants received monetary compensation for participation (24€). Due to technical issues, behavioral and physiological data from 9 participants (ID 1-9) were lost for the recognition phase and questionnaire measures are missing from 5 participants (ID 1-2, 4, 22 and 30). Six participants (ID 8 and 21 for encoding, and ID 21, 27, 44, 47 and 50 for recognition) were excluded from startle analysis due to poor data quality (i.e., excessive EMG artifacts) and two from SCR analyses (ID 8 and 30) due to non-responding. All participants provided informed written consent to the experimental protocol which was approved by the ethics committee of the Medical Faculty Mannheim, Heidelberg University (Germany) and complies with the APA ethical standards and the Declaration of Helsinki.

Sample size was determined based on previous studies using a similar threat-of-shock procedure (e.g., [25] – [26]) and estimation of sample size using G*Power analyses [27] assuming a medium effects size (f = .2), power (1-ß = .9), and correlation among repeated measures (r = .05) suggesting N = 48 to find reliable effects regarding item/context effects for the main startle analyses.

Materials and task

Sixty faces (30 female) with neutral facial expressions were selected from the Karolinska Directed Emotional Faces (KDEF; [28]) and the Radboud Faces Database (RaFD; [29]). The pictures were processed with a photo editor (Adobe Photoshop CS2) and cropped to the same size (442×606 pixels) to reduce differences between the datasets. The stimulus set was randomly divided into three subsets of 20 face actors each (Set A, B, and C). During an encoding session, pictures of Set A and B were presented in a random order. Specifically, Set A was presented with a specific colored background frame (e.g., blue; RGB values: 0,255,0), whereas Set B was presented with a different colored frame (e.g., green; RGB values: 0,0,255; 1280×1024 pixels; see Figure 1). Assignment of subsets to background color was counterbalanced across participants. Face-context compounds were presented in a trial-to-trial pseudorandom order with the restriction of no more than two repetitions of the same background color. Pictures were presented for 6 s with a flexible ITI of 8-12 s. To provoke the defensive startle reflex, 10 auditory startle probes (104 dB, white noise) were administered per picture set, with a variable onset at 4.5-5.5s after picture onset (3 startle during ITI). To examine incidental learning and in order to keep recognition performance comparably low to our previous studies ([6], [5]), the recognition task was not mentioned to the participants before the encoding session.

During the recognition session, all pictures (old Sets A and B, and new Set C) were presented intermixed without background colors and a combined item/source memory task was performed. Participants indicated by button press whether the face was previously seen (old) and if so in which context, or whether the face was new. They were instructed to guess the context if they remembered the face but not the background color, instead of choosing new. In total, 30 startle probes were equally distributed to the conditions old-threat, old-safe, and new faces; 3 additional probes were presented during ITI.

Stimuli were presented on a 22-inch computer screen placed approximately 0.8m in front of the participants using Presentation software (Version 20.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

Procedure

Participants completed the questionnaires on general and social anxiety as well as trauma history one to two days before their lab appointment, where the state version of the STAI was administered a second time. They were then seated in the experimental room and the electrodes for the physiological measures were attached. Following, participants rated the two background colors regarding valence and arousal using the Self-Assessment Manikin (SAM; [30], on a scale from 1-9) as well as perceived threat using a Likert scale (from *not at all* 0 to 10 *highly threatening*) which served as a baseline rating. Afterwards a habituation phase served to familiarize participants with the auditory startle probes (8 trials, not analyzed).

Next, a (fake) shock electrode was placed on the non-dominant forearm. The participants were then verbally instructed that they could receive up to three maximal unpleasant but not yet painful electric shocks while one background frame was present (e.g., blue serves as threat context) but not while the other was present (e.g., green serves as safety context). Verbal threat-of-shock instructions were used to investigate aversive expectation instead of direct experience [31]. Color assignment to threat and safety was counterbalanced across participants. In the following encoding session, participants were instructed to attentively watch all presented face. After encoding, the shock electrode was detached, and participants rated the context colors regarding valence, arousal, and perceived threat as a follow-up measure. The recognition task followed immediately. Afterwards participants completed a short interview and were debriefed.

Data recording and reduction

Electrodermal activity (EDA), and electromyography (EMG) were recorded (1000Hz sampling rate) as indicators for autonomic nervous system activation using Ag-AgCl electrodes, AcqKnowledge software and Biopac amplifiers (BIOPAC Systems; Golet, CA). Additionally, heart rate (ECG) was assessed, but not analyzed. For skin conductance, electrodes were placed at the hypothenar eminence of the non-dominant hand. Skin conductance data was down sampled to 20Hz offline and noise was attenuated using Butterworth Zero Phase 2Hz low- and a 0.05 Hz high-pass filter. For the startle EMG, two electrodes were attached below the right eyelid, measuring the electromyogram of the orbicularis oculi muscle [32]. During recording frequencies below 28Hz and above 500Hz were filtered out with a bandpass filter (24 dB/octave roll-off) and a 50Hz Notch filter was applied to remove noise from the power line offline. This data was then rectified and smoothed with a moving average procedure (50 ms) in BrainVision Analyzer 2.0 (BrainProducts, Munich, Germany).

Startle responses were scored with a semi-automated procedure as maximum peak in the 21 – 150 ms time window following each startle probe. Peak amplitude was calculated relative to a mean baseline period (50 ms preceding startle response time window; cf. [33]). Skin conductance response (SCR) to startle probe and picture onsets were calculated as the maximum increase in skin conductance in the interval of 1 to 4.5 s (relative to a 2 s pre-stimulus period). A minimum threshold of 0.02 μS was used for zero-response detection, and range and distribution correction were applied.

Data analysis

Self-report data

Valence, arousal, and threat ratings were submitted to a repeated measure ANOVA with Time (T0 baseline vs. T1 follow-up) and Context (threat vs. safety) as within-subjects factors. Helmert contrasts were used to follow-up on main effects at a significance level of $p < .05$. Greenhouse-Geisser corrections were applied where necessary and as a measure of effect size, partial eta squared (η_p^2) is reported.

Behavioral data

Regarding recognition performance, hierarchical multinomial processing tree modeling (MPT; [34]) was used providing parameter estimates for item- and source-memory as well as guessing parameters. Moreover, this approach enables examining co-variations between memory parameters and inter-individual questionnaire measures like anxiety and depression scores. The two-high-threshold source monitoring model (2HTSM, [35]) for two sources was used.

Three discrete answer categories (threat context “T”, safe context “S”, and new “N”) are assumed to be derived from certain latent cognitive processes, modelled as probabilities in three decision trees (see Figure 2). Specifically, item recognition (D_T, D_S & D_N), source recognition (d_t, d_s), item guessing (b), and source guessing (g). The first decision tree represents the case of an item presented with a threatening context. With probability D_T the item is recognized as being previously seen (i.e., old) and with the probability d_t the source (i.e., threatening context) is also recognized. If the item is not recognized as old ($1-D_T$), participants correctly guess with probability b that the item was old. In that case, the source must be guessed as well, and with probability g the correct source is chosen (i.e., threat). The second decision tree represents the case of an item previously presented with a safe context and the probabilities derive accordingly. The third decision tree represents a previously unseen (i.e., new) item. With probability D_N the item is recognized correctly as being new, with probability $1-D_N$ it is falsely classified as old and the parameters arise accordingly to the first two decision trees. In this current form, the model is not identifiable. Therefore, the parameters for recognizing old safe faces and new faces were restricted to be equal ($D_S = D_N$), while D_T remained unrestricted. These restrictions are typical of the 2HTSM (e.g., [35] – [36]). Additionally, guessing parameters were set equal (b = g), under the assumption that guessing tendencies would be the same regardless of item recognition [36] – [37]. Analyses were conducted using the TreeBUGS package [34] for R [38], questionnaire data were entered as covariates into the model.

Additionally, hit rates (HR), false alarm rates (FAR), item recognition (HR – FAR) and the conditional source identification measure (CSIM, number of correctly identified source attributions divided by the total number of targets, [5]) were calculated and submitted to an ANOVA with the factor Context (threat vs. safety).

Physiological data

Startle EMG, SCR, and HR were submitted to repeated measures ANOVAs with the factor Context (threat vs. safety), separately for the encoding and recognition phase. For SCR the factor of Block (first half vs. second half) was included to account for habituation effects.

In order to explore the probability of null effects of our central hypotheses, we conducted Bayesian analyses [39], using the R based software package JASP 0.16.1 [40]. Thereby we focused on the central effect of Context on behavioral performance, startle potentiation, and SCR. Additionally, for SCR the factor Block and the interaction Context × Block were included. Using Monte-Carlo sampling 10,000 iterations and default prior scaling factors (for fixed effects = 0.5, random effects = 1, r covariates = 0.354, Bayes factors (BF) were estimated [41]. BF inclusion scores (BF_{incl}) are reported that inform about how much the inclusion of the factor (e.g., Context) is supported by the data, compared to the null-model. A value below 1 (above 1) indicated that the data is more likely for the null-hypotheses than for the alternative hypotheses (and vice versa).

Greenhouse-Geisser corrections were applied where necessary, and the partial eta square (η_p^2) is reported as a measure of effect size. To control for type 1 error, Bonferroni correction was applied for post hoc t tests.

Results

Self-report data

The rating data revealed an overall successful subjective induction of threat anticipation (see Figure 3). For the valence ratings, neither the two contexts nor the two time points T0 (baseline) and T1 (follow-up) were rated differently, Time $F(1,47) = .38$, $p = .54$, $\eta_p^2 = .01$, Context $F(1,47) = 1.35$, $p = .25$, $\eta_p^2 = .03$. A significant Time × Context interaction showed, however, that the safety context was perceived as more pleasant than the threat context after threat induction but not at baseline, Time × Context $F(1,47) = 29.39$, $p < .001$, $\eta_p^2 = .39$. The overall level of arousal was higher during T1 than during T0, Time $F(1,47) = 10.03$, $p < 0.01$, $\eta_p^2 = .18$. While in general the contexts did not differ, Context $F(1,47) = 2.32$, $p = .13$, $\eta_p^2 = .05$, post-hoc tests revealed that, after the threat induction, the threat relative to safety context was rated more arousing, Time × Context $F(1,47) = 27.18$, $p < .001$, $\eta_p^2 = .37$. Threat levels increased over time, Time $F(1,47) = 20.98$, $p < .001$, $\eta_p^2 = .31$ and in intensity, Context $F(1,48) = 4.10$, $p = .05$, $\eta_p^2 = .08$, with post-hoc tests showing that the threat context was perceived as more threatening than the safe context only during T0 and not during T1, Time × Context $F(1,47) = 26.22$, $p < .001$, $\eta_p^2 = .36$. There were no covariation effects with questionnaire scores on any of these measures.

Behavioral data

Table 1 contains the parameter estimates of the hierarchical MPT. Using Bayesian modelling, a priori distribution of the parameters is used before analyses and compared to a posterior distribution which is based on the data using Baye's theorem. The Bayesian confidence intervals (BCIs) explains within which range the true parameter lies in the posterior distribution, given data and prior distribution. Overlapping BCIs indicate non-significant differences between parameter estimates. While participants were moderately able to recognize items as old or new, it did not matter whether they had seen them in a threatening or a safe context during encoding. This is indicated by the overlapping BCIs of the D_T and D_S parameters.

Moreover, source recognition was poor ($d_T = .04$ and $d_S = .01$ with large Bayesian confidence intervals for the threatening source [.00 - .39]) and did not differ between a threatening or safe source from encoding, evident in overlapping BCIs of the d_t and d_s parameters. The guessing parameters reflected the probabilities of the events for source guessing (g = .48 with half of the faces from a threatening and half of the faces from a safe context). The probability of a face having been presented before during encoding compared to new was 2:1. There was a slight conservative guessing tendency of classifying an old face as a face being new (b = .58 with the actual probability of a face being old of .67). No significant associations emerged between parameter estimates and questionnaire scores, which were entered as covariates into the model. Model fit (Klauer's [42] test statistics $T1$ p = .25) was good.

Using conventional recognition parameters, mean hit rates and false alarm rates for the threatening context were HR $M_{threat} = .76$ (SD = .15), FAR $M_{threat} = .20$ (SD = .17) and for the safe context

HR M_{safe} = .73 (SD = .16) and M_{safe} = .20 (SD = .15). Item recognition (HR − FAR) did not differ based on Context, $F < 1$, $p = .40$, $\eta_p^2 = .02$, $BF_{incl} = 0.305$. Likewise, source identification did not differ significantly as a function of context conditions, Context $F < 1$, $p = .50$, $\eta_p^2 = .01$, $BF_{incl} = 0.290$ (M_{threat} = .21 [SD = .07], M_{safe} = .22 [SD = .05]). This means that it is 3.28 and 3.45 times more likely that there is no difference between the two context conditions for both behavioral measures. None of the questionnaire measures had an impact on memory performance, *Fs* < 1, *ps* > .73.

Table 1. Mean parameter estimates of the latent-trait MPT model for the recognition performance of the item/source memory task.

Parameter	M [95 % BCI]	SD
g	0.48 [0.45 – 0.52]	0.02
b	0.58 [0.49 – 0.67]	0.05
d_T	0.04 [0.00 – 0.39]	0.11
D_T	0.28 [0.05 – 0.48]	0.11
d_S	0.01 [0.00 – 0.04]	0.02
$D_S = D_N$	0.41 [0.32 – 0.49]	0.04

Note. For the group-level estimates, posterior means (and SDs) are shown. BCI = Bayesian confidence interval. DT, DS and DN =face recognition parameters, dT, dS = context memory parameters and b, g = guessing probabilities.

Startle reflex

As expected for the encoding phase, the startle reflex was potentiated for faces presented with the threatening compared to the safe background, Context $F(1,45) = 37.20$, $p < .001$, $\eta_p^2 = .45$ (see Figure 4A). A marginally significant Context × Depression interaction indicated that the threat-potentiated startle was more pronounced for more depressed participants, $F(1, 41) = 3.76$, $p = .06$, $\eta_p^2 = .08$. No significant interaction emerged for other questionnaire measures.

Regarding the recognition phase, no difference in startle potentiation was found, Context $F(2,70) < 1$, $p = .64$, $\eta_p^2 = .01$, $BF_{incl} = 0.133$, making the null hypotheses 7.52 times more likely than the alternative hypothesis. Individual differences based on the questionnaire scores did not modulate the findings.

Skin conductance responses to startle probes

Skin conductance responses to startle probes were more pronounced during the first relative to the second half of the encoding session, Block $F(1,43) = 31{,}43)$, $p < .001$, $\eta_p^2 = .42$. No main effect of Context emerged, $F < 1$, $p = .53$, $\eta_p^2 = .01$, however a significant interaction Context × Block, $F(1,43) = 6.56$, $p < .05$, $\eta_p^2 = .13$. Post-hoc tests revealed that the threatening context elicited a more pronounced SCR compared to the safe context only during the second half of the encoding session, $F(1,43) = 6.83$, $p < .05$, $\eta_p^2 = .14$. Moreover, (marginally) significant interactions emerged for Context with depression, trait-anxiety, and social anxiety, $Fs(1, 40) = 4.01$, 4.93, and 6.02, $ps = .05$, and $< .05$, $\eta_p^2 = .09$, .11, and .13. These interactions indicate more pronounced differentiation between threat and safety context with higher depression and anxiety scores.

For the recognition phase, startle locked SCRs were not modulated by Context, $F < 1$, $p = .45$, $\eta_p^2 = .03$, $BF_{incl} = 0.074$, or Block $F(1,28) = 3.21$, $p = .08$, $\eta_p^2 = .10$, $BF_{incl} = 0.995$. There was also no significant Context × Block interaction, $F < 1$, $p = .45$, $\eta_p^2 = .03$, $BF_{incl} = 0.04$, making the respective null hypotheses 13.51, 1.01, and 25 times more likely than the alternative hypotheses.

Skin conductance responses to picture onset

During the encoding session, SCRs locked to picture onset varied as a function of Context, $F(1,43) = 29.93$, $p < .001$, $\eta_p^2 = .41$, and Block, $F(1,43) = 60.55$, $p < .001$, $\eta_p^2 = .59$, indicating more pronounced SCRs during threat compared to safety, and during the first compared to the second half of the encoding session (see Figure 4B). There was no significant interaction Context × Block, $F(1,43) = 2.10$, $p = .16$, $\eta_p^2 = .05$, nor any interactions with the questionnaire measures, $Fs < 1$, $ps > .43$.

During recognition, SCRs elicited did also not vary between threat and safety, Context, $F(2,72) = 1.05$, $p = .35$, $\eta_p^2 = .03$, $BF_{incl} = 0.067$, nor between first and second phase of the recognition session, Block $F < 1$, $p = .83$, $\eta_p^2 = .00$, $BF_{incl} = 0.106$. The Context × Block interaction was not significant, $F(2,72) = 1.53$, $p = .22$, $\eta_p^2 = .04$, $BF_{incl} = 0.011$. Thus, the respective null hypotheses were 14.93, 9.43, and 90.91 times more likely than the alternative hypothesis. Taken together, in the recognition phase, there was no indication that autonomic arousal was enhanced for faces that had been previously presented with a threatening compared to a safety context. There were no significant influences of the questionnaire measures, $Fs < 1$, $ps > .51$.

Discussion

Previous research has shown that encoding unfamiliar faces within an aversive anticipatory context (i.e., threat of shock) leads to later threat-selective neural processing (indicated by event-related brain activity), although recognition performance was low [6], [9]. In the present study, we examined the hypotheses whether autonomic arousal can be found for such threat-associated but non-recognized unfamiliar face pictures. Psychophysiological measures of the startle reflex and SCR did not support this hypothesis.

Aversive apprehensions were successfully induced via verbal threat instruction in the encoding session and resulted in a more unpleasant, arousing, and threatening rating of the threat compared to the safety context. Moreover, similar to previous research, a threatening context activated the autonomic and somatic nervous system during the encoding session, as evidenced by threat-enhanced skin conductance responses and potentiated startle reflexes compared to the safety context (e.g., [43]). As expected for an implicit learning task, memory for unfamiliar faces was rather poor and did not vary as a function of contextual threat or safety [6], [5]. In fact, participants were barely able to recognize whether they had seen a face before, and thus were unable to identify the contextual situation (i.e., threat or safety source) of a previous encounter. Similarly, autonomic and somatic activation did not vary during the recognition session, neither for faces from a threat or safety source, nor for newly presented faces. Moreover, different measures of psychopathology (i.e., depression, social and state/trait anxiety, early maltreatment) did not affect psychophysiological responding to faces during the recognition session. These findings are further supported by Bayesian analyses, showing that the null hypotheses in the recognition phase (i.e., no difference between faces from a threat or safety context or new faces) were much more likely than the alternative hypotheses.

A dangerous situation triggers the activation of defensive psychophysiological systems in order to avoid or, if a flight reaction is not possible, at least minimize harm [44]. Such response programs to danger cues include the activation of the autonomic and somatic nervous system and can persist even when the dangerous situation has evidently passed. This is particularly true when the threat was learned through social means such as verbal instructions [45]–[46]. The pure anticipation of threatening events (e.g., unpleasant electric shocks; [47]) is sufficient to provoke a relatively persistent state of aversive apprehension even when the threatening situation never occurred (e.g., across repeated test days without shock reinforcement; [33], [43]).

However, once the threat has been averted and its signals have disappeared, a conscious memory of people and situations associated with the threat seems necessary to be better prepared for future encounters [48]. While we did not find support for this hypothesis in the present study design, several aspects and alternative hypotheses need to be considered. First, the removal of the shock electrode after the encoding phase together with the instruction that no electrical shock can follow after this point may have terminated the real threat (instructed extinction; [49]). Second, the complete absence of threat cues during recognition probably triggered a general downregulation of defensive behaviors [50]. However, compared to the startle reflex activity during a safe encoding context, no reduction or downregulation was observed in the recognition session for formerly threatening, safe or new faces. Third, aversive apprehensions have been suggested to positively improve cognitive performance [18]. For example, arousal improves accuracy in a sustained attention to a response task [51] or in recognizing neutral objects from emotional backgrounds [19]. In another study, memory of context (arousing vs. non-arousing) was found to be impaired by the presence of negative information, whereas threat of shock enhanced item memory [52]. Therefore, negative affect appears to impair memory for associations, whereas storage of negative perceptual representation is spared or even enhanced. Fourth, in the present study, the facial stimuli were prominently placed in the foreground, while contextual threat and safety colors served as a background. Although participants were instructed to attend to every picture, the colored background indicated threat of electrical shocks or safety and was relevant for adaptive behavioral response priming (e.g., avoidance, heightened attention). In contrast, faces were a prerequisite for source identification during the recognition session [36]. Together with a possible item-source dissociation during encoding, this might have led to an overall decrease in psychophysiological responding to a potentially dangerous situation [6].

Although poor memory performance was intended in this study (to investigate the possible activation of defensive systems without conscious recollection), future studies should strengthen the item-source association during encoding to promote familiarity-based retrieval [53]–[54]. Moreover, future studies are needed to investigate the effect of threat on recognized faces by increasing memory performance, for example by lowering the number of faces to be remembered. Here, explicit learning instructions and increasing the number of presentations per face could be pertinent. This procedure could be used to clarify whether defensive reactions are triggered even when there is no conscious recognition of previous threatening encounters.

Although the recognition of unfamiliar faces was expectedly poor [2], and the association of the faces with the threatening (and safe) context(s) not explicitly emphasized and strengthened, this study aligns with other findings that show the impact of contextual surroundings on the perception and processing of otherwise neutral faces [55]–[56], [19], [6]. Extending these findings, we found enhanced SCRs to the onset of faces (and startle probes) within a threat compared to safety context, indicating enhanced arousal and alertness to these faces–context compounds. This effect diminished, when eliminating the threat-related part of the face-context associations during recognition, where SCRs were similar for all faces (i.e., threat, safe, and new), especially when being unable to remember the presented stimuli. This underscores the importance of the interplay of bottom-up (automatic) and top-down (expectation-based) factors in threat perception [57]–[59]. It also supports the claim that neutral facial expressions do not convey sufficient safety information and are disregarded in a dangerous context [43]. In addition, similar contexts during encoding and recognition has long been found to boost memory performance [60]–[61]. As the context conditions fundamentally changed from encoding to recognition, a cued recognition task with presenting the faces together with all contexts from encoding could have led to a renewed autonomic nervous system activation.

From a clinical perspective, the present study adds insight to several important points. While previous studies have shown that a threatening environment can change perception, attention, and memory processes (e.g., [6], [9]), here we found no indication for our key prediction that threat-encoded faces will lead to enhanced defense activation during recognition. However, psychophysiological responding to threat/safety situations varied with interindividual differences in depression and trait-anxiety measures during encoding of face–context compounds. Specifically, threat-enhanced skin conductance response increased with anxiety and depression scores. This differentiation did not persist into the recognition phase of the experiment, possibly due to response habituation. Alternatively, these findings are in line with previous research showing that anxiety, at least at low-moderate levels, has no significant impact on implicit memory and recognition and does not favor a memory bias towards threatening information [62]. Participants with moderate-high levels of anxiety and depression should be included in future studies to gain insight to the effect of former contextual arousal on face recognition.

In summary, a first encounter with an unfamiliar person that was combined with contextual threat leads to activation of defensive response systems (threat-potentiated startle reflex and SCRs during encoding). However, this effect disappeared during the second encounter with this person without contextual information (recognition session), indicating that the presence of contextual threat signals is necessary to trigger defensive responding, especially when memory of the threat–face association is poor.

Declarations

Data availability statement

The datasets generated during and/or analyzed during the current study are available in the OSF repository, https://osf.io/8afpm/?view_only=57d8e169e60a42e7b5d2ecfebff57170

Acknowledgements

We are grateful to Maike Haag, Kathrin Zauner and Niclas Willscheid for data collection. This research was supported by the German Research Foundation (DFG) grant to Florian Bublatzky (BU 3255/1-2) and a scholarship by the Studienstiftung des Deutschen Volkes awarded to Sabine Schellhaas.

Declaration of competing interests

The authors report there are no competing interests to declare.

Author contributions

S.S. was involved in study design, supervised data collection and analyses, wrote and revised the manuscript; F.B. designed the study, supervised data analyses, revised the manuscript, and acquired funding; C.S. was involved in study design and revised the manuscript.

References

1. Davies, G., & Milne, A. (1982). Recognizing faces in and out of context. Current psychology, *2*(1), 235–246. https://doi.org/10.1007/BF03186766
2. Burton, A.M., & Jenkins, R. (2011). *Unfamiliar face perception*. The Oxford handbook of face perception, 28, 287–306.
3. Young, A. W., & Burton, A. M. (2017). Recognizing Faces. Current Directions in Psychological Science, *26*(3), 212–217. https://doi.org/10.1177/0963721416688114
4. Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychological Review, *108*(3), 483–522. https://doi.org/10.1037/0033-

295X.108.3.483

5. Arnold, N. R., González Cruz, H., Schellhaas, S., & Bublatzky, F. (2021). A multinomial modelling approach to face identity recognition during instructed threat. Cognition and Emotion, *35*(7), 1302–1319. https://doi.org/10.1080/02699931.2021.1951175
6. 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
7. Carretié, L., Albert, J., López-Martín, S., & Tapia, M. (2009). Negative brain: an integrative review on the neural processes activated by unpleasant stimuli. International Journal of Psychophysiology, *71*(1), 57–63. https://doi.org/10.1016/j.ijpsycho.2008.07.006
8. Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. Behaviour research and therapy, *38*(4), 319–345. https://doi.org/10.1016/S0005-7967(99)00123-0
9. Schellhaas, S., Schmahl, C, & Bublatzky, F. (under review). *Social learning in individuals with adverse childhood experiences: Memory processes as a function of instructional and observational threat and safety learning* [Manuscript submitted for publication].
10. Shackman, A. J., Stockbridge, M. D., Tillman, R. M., Kaplan, C. M., Tromp, D. P. M., Fox, A. S., & Gamer, M. (2016). The Neurobiology of Dispositional Negativity and Attentional Biases to Threat: Implications for Understanding Anxiety Disorders in Adults and Youth. Journal of Experimental Psychopathology, *7*(3), 311–342. https://doi.org/10.5127/jep.054015
11. Robinson, O. J., Charney, D. R., Overstreet, C., Vytal, K., & Grillon, C. (2012). The adaptive threat bias in anxiety: Amygdala–dorsomedial prefrontal cortex coupling and aversive amplification. NeuroImage, *60*(1), 523–529. https://doi.org/10.1016/j.neuroimage.2011.11.096
12. Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon, C. (2008). Generalization of conditioned fear-potentiated startle in humans: Experimental validation and clinical relevance. Behaviour Research and Therapy, *46*(5), 678–687. https://doi.org/10.1016/j.brat.2008.02.005
13. McTeague, L. M., & Lang, P. J. (2012). The anxiety spectrum and the reflex physiology of defense: from circumscribed fear to broad distress. Depression and anxiety, *29*(4), 264–281. https://doi.org/10.1002/da.21891
14. Lang, P. J., & McTeague, L. M. (2009). The anxiety disorder spectrum: Fear imagery, physiological reactivity, and differential diagnosis. Anxiety, Stress, & Coping, *22*(1), 5–25. https://doi.org/10.1080/10615800802478247
15. Grillon, C., & Charney, D. R. (2011). In the face of fear: Anxiety sensitizes defensive responses to fearful faces: Fearful faces are not feared. Psychophysiology, *48*(12), 1745–1752. https://doi.org/10.1111/j.1469-8986.2011.01268.x
16. Bublatzky, F., Guerra, P. M., Pastor, M. C., Schupp, H. T., & Vila, J. (2013). Additive effects of threat-of-shock and picture valence on startle reflex modulation. PloS ONE, *8*, e54003. https://doi.org/10.1371/journal.pone.0054003

17. Lobo, I., Portugal, L. C., Figueira, I., Volchan, E., David, I., Garcia Pereira, M., & de Oliveira, L. (2015). EEG correlates of the severity of posttraumatic stress symptoms: A systematic review of the dimensional PTSD literature. Journal of Affective Disorders, *183*, 210–220. https://doi.org/10.1016/j.jad.2015.05.015
18. Dolcos, F., Denkova, E., & Dolcos, S. (2012). Neural correlates of emotional memories: A review of evidence from brain imaging studies. Psychologia, *55*(2), 80–111. https://doi.org/10.2117/psysoc.2012.80
19. Ventura-Bort, C., Löw, A., Wendt, J., Dolcos, F., Hamm, A. O., & Weymar, M. (2016a). 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
20. Wangelin, B. C., Bradley, M. M., Kastner, A., & Lang, P. J. (2012). Affective engagement for facial expressions and emotional scenes: The influence of social anxiety. Biological Psychology, *91*(1), 103–110. https://doi.org/10.1016/j.biopsycho.2012.05.002
21. Stangier, U., & Steffens, M. (2001). Social Phobia Inventory—deutsche Fassung [German ver-sion]. Psychologisches Institut der Universität Frankfurt am Main.
22. Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). Das State-Trait-Angstinventar [The state-trait anxiety inventory]. Beltz Test GmbH.
23. Hautzinger, M., Keller, F., & Kühner, C. (2006). BDI-II Beck Depressions-Inventar [BDI-II Beck depression inventory]. Harcourt Test Services.
24. Bernstein, D.P., Fink, L., Handelsman, L., & Foote, J. (1998). Childhood trauma questionnaire. Assessment of family violence: A handbook for researchers and practitioners. https://psycnet.apa.org/doi/10.1037/t02080-000
25. Bradley, M. M., Zlatar, Z. Z., & Lang, P. J. (2018). Startle reflex modulation during threat of shock and “threat” of reward. Psychophysiology, *55*(2), e12989. https://doi.org/10.1111/psyp.12989
26. Dunning, J. P., DelDonno, S., & Hajcak, G. (2013). The effects of contextual threat and anxiety on affective startle modulation. Biological Psychology, *94*(1), 130–135. https://doi.org/10.1016/j.biopsycho.2013.05.013
27. Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, *39*, 175–191. https://doi.org/10.3758/BF03193146
28. Lundqvist, D., Flykt, A., & Öhman, A. (1998). Karolinska directed emotional faces [Database of standardized facial images]. Department of Clinical Neuroscience, Psychology section. Karolinska Institutet, 91, 630.
29. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., & Van Knippenberg, A.D. (2010). Presentation and validation of the Radboud Faces Database. Cognition and Emotion, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076.
30. Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry, *25*(1), 49–59.

https://doi.org/10.1016/0005-7916(94)90063-9.

31. Olsson, A., & Phelps, E. A. (2004). Learned fear of “unseen” faces after Pavlovian, observational, and instructed fear. Psychological science, *15*(12), 822–828. https://doi.org/10.1111/j.0956-7976.2004.00762.x
32. Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. Psychophysiology, *42*(1), 1–15. https://doi.org/10.1111/j.1469-8986.2005.00271.x
33. Bublatzky, F., Gerdes, A.B., & Alpers, G.W. (2014). The persistence of socially instructed threat: Two threat-of-shock studies. Psychophysiology, *51*(10), 1005–1014. https://doi.org/10.1111/psyp.12251
34. Heck, D.W., Arnold, N.R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. Behavior Research Methods, *50*, 264–284. https://doi.org/10.3758/s13428-017-0869-7
35. Bayen, U.J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. Journal of Experimental Psychology: Learning, Memory, and Cognition, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197.
36. Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old–new recognition. Emotion, *17*(1), 120. https://doi.org/10.1037/emo0000211
37. Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. Psychonomic Bulletin & Review, *20*(2), 326–333. https://doi.org/10.3758/s13423-012-0342-7
38. Core Team, R. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/
39. Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, *90*(430), 773–795.
40. JASP Team (2022). JASP (Version 0.16.1)[Computer software].
41. Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. Journal of mathematical psychology, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001
42. Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. Psychometrika, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0
43. Bublatzky, F., Schellhaas, S., & Guerra, P. (2022). The mere sight of loved ones does not inhibit psychophysiological defense mechanisms when threatened. Scientific reports, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-06514-y
44. Gilbert, P. (2001). Evolutionary approaches to psychopathology: The role of natural defences. Australian & New Zealand Journal of Psychiatry, *35*(1), 17–27. https://doi.org/10.1046/j.1440-1614.2001.00856.x

45. Mertens, G., & De Houwer, J. (2016). Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. Biological psychology, *113*, 91–99. https://doi.org/10.1016/j.biopsycho.2015.11.014
46. Mertens, G., Boddez, Y., Sevenster, D., Engelhard, I. M., & De Houwer, J. (2018). A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions. Biological Psychology, *137*, 49–64. https://doi.org/10.1016/j.biopsycho.2018.07.002
47. Grillon, C., Ameli, R., Woods, S. W., Merikangas, K., & Davis, M. (1991). Fear-Potentiated Startle in Humans: Effects of Anticipatory Anxiety on the Acoustic Blink Reflex. Psychophysiology, *28*(5), 588–595. https://doi.org/10.1111/j.1469-8986.1991.tb01999.x
48. Lang, P. J., Davis, M., & Öhman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. Journal of affective disorders, *61*(3), 137–159. https://doi.org/10.1016/S0165-0327(00)00343-8
49. Luck, C. C., & Lipp, O. V. (2016). Instructed extinction in human fear conditioning: History, recent developments, and future directions. Australian Journal of Psychology, *68*(3), 209–227. https://doi.org/10.1111/ajpy.12135
50. Rowles, M. E., Lipp, O. V., & Mallan, K. M. (2012). On the resistance to extinction of fear conditioned to angry faces. Psychophysiology, *49*(3), 375–380. https://doi.org/10.1111/j.1469-8986.2011.01308.x
51. Aylward, J., Valton, V., Goer, F., Mkrtchian, A., Lally, N., Peters, S., Limbachya, T., & Robinson, O. J. (2017). The impact of induced anxiety on affective response inhibition. Royal Society Open Science, *4*(6), 170084. https://doi.org/10.1098/rsos.170084
52. Bisby, J. A., & Burgess, N. (2014). Negative affect impairs associative memory but not item memory. Learning & Memory, *21*(1), 21–27. https://doi.org/10.1101/lm.032409.113
53. Diana, R. A., Van den Boom, W., Yonelinas, A. P., & Ranganath, C. (2011). ERP correlates of source memory: Unitized source information increases familiarity-based retrieval. Brain research, *1367*, 278–286. https://doi.org/10.1016/j.brainres.2010.10.030
54. Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. Perspectives on psychological science, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234
55. 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
56. Klein, F., Iffland, B., Schindler, S., Wabnitz, P., & Neuner, F. (2015). This person is saying bad things about you: the influence of physically and socially threatening context information on the processing of inherently neutral faces. Cognitive, Affective, & Behavioral Neuroscience, *15*(4), 736–748. https://doi.org/10.3758/s13415-015-0361-8
57. Sussman, T. J., Jin, J., & Mohanty, A. (2016). Top-down and bottom-up factors in threat-related perception and attention in anxiety. Biological psychology, *121*, 160–172. https://doi.org/10.1016/j.biopsycho.2016.08.006

58. 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
59. Bublatzky, F., Schellhaas, S., & Paret, C. (under review). *Comparing apples and oranges: Electrocortical correlates of decision-making and reward reversal learning under threat* [Manuscript submitted for publication].
60. Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological review, *80*(5), 352. https://doi.org/10.1037/h0020071
61. Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: the ICE theory. Journal of Experimental Psychology: General, *128*(4), 403. https://doi.org/10.1037/0096-3445.128.4.403
62. Mitte, K. (2008). Memory bias for threatening information in anxiety and anxiety disorders: a meta-analytic review. Psychological bulletin, *134*(6), 886. https://doi.org/10.1037/a0013343

Figures

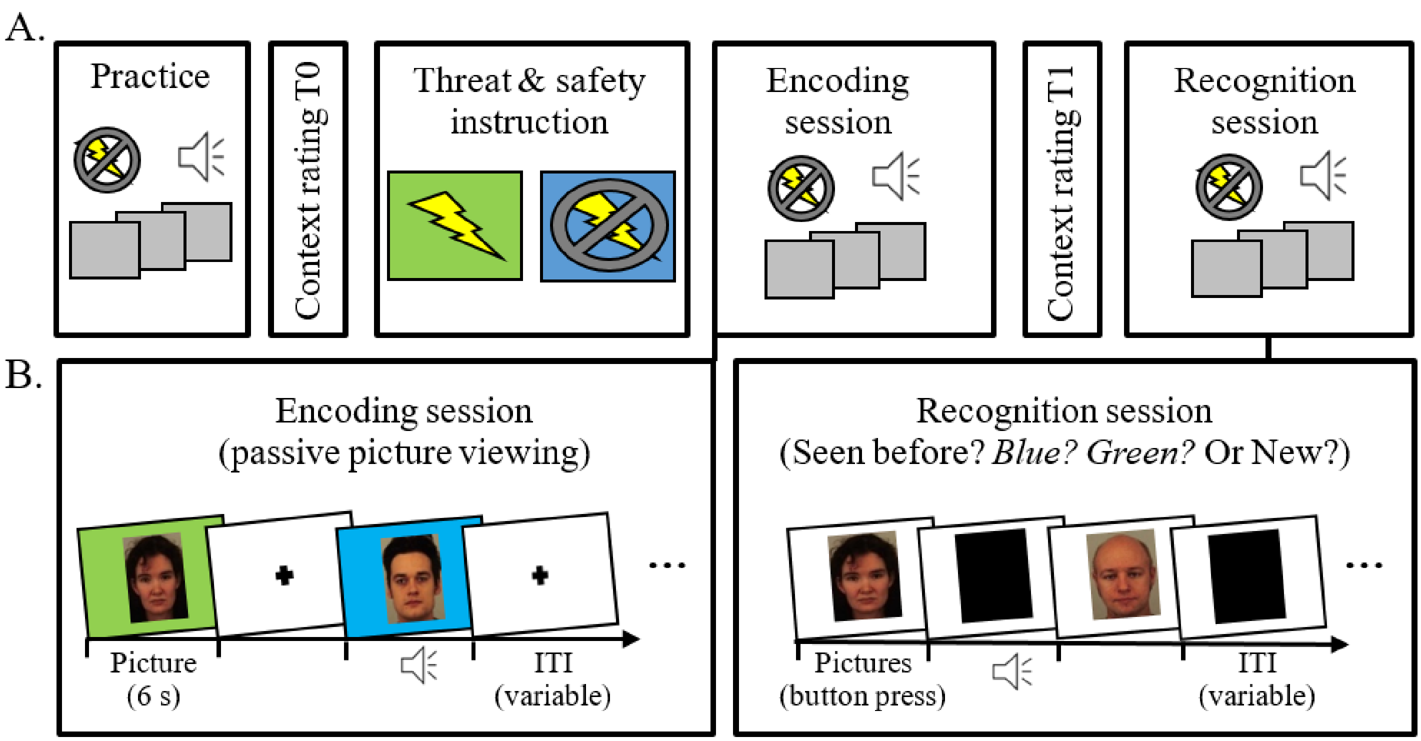

Figure 1

Schematic illustration of the experimental procedure. (A) Following initial practice trials, context colors were rated regarding valence, arousal and threat (T0) and then threat-of-shock and safety based on the colors were verbally instructed. Afterwards, an encoding session started, after which the context colors were rated a second time (T1). Finally, participants performed a recognition session. (B) During the encoding session 40 faces with neutral facial expressions were presented in front of either of the two alternating background colors (i.e., threat and safety) and the participants were instructed to view each face attentively. A startle sound (white noise) was presented for half of the trials. In the recognition session participants saw the 40 faces from encoding and 20 new ones and had to indicate by button press in front of which color they had been presented during encoding or whether they were new. Again, half of the trials were accompanied by a startle sound.

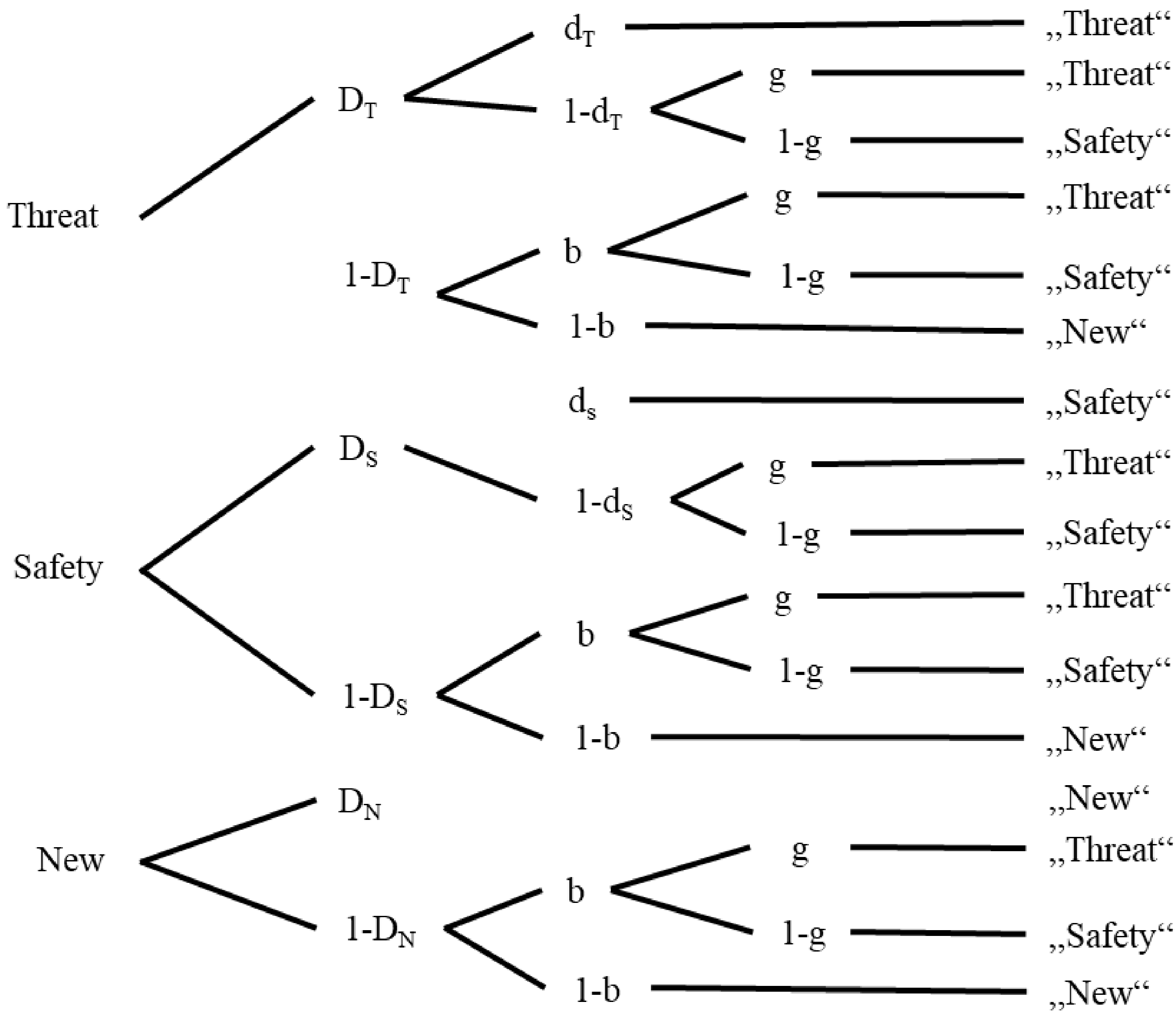

Figure 2

Submodel 5d of the two-high-threshold model of source monitoring (Adapted from Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; originally by Bayen et al., 1996). The model parameters represent probabilities; DT = probability of detecting that a face is from encoding session in a threatening context; DS = probability of detecting that a face is from encoding session in a safe context; DN = probability of detecting that a face is new; dT = probability of correctly remembering the threat context of a face; dS = probability of correctly remembering the safe context of a face; g = probability of guessing that a face is from the threat context; b = probability of guessing that a face is from the encoding session.

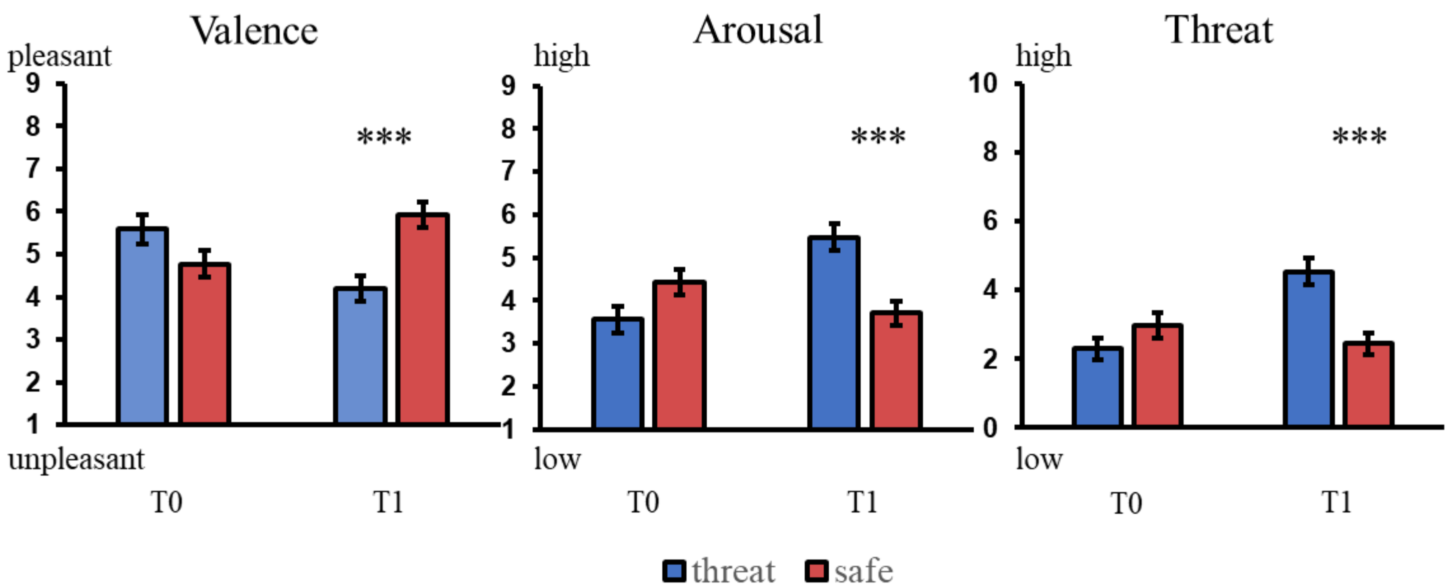

Figure 3

Mean valence, arousal (scale of 1 to 9) and threat (scale of 0 to 10) ratings of the instructed threat and safety contexts, indicated by the colors blue and green (random assignment) before the threat instruction (T0) and after the encoding session (T1) (M and SEM, *** p < .001).

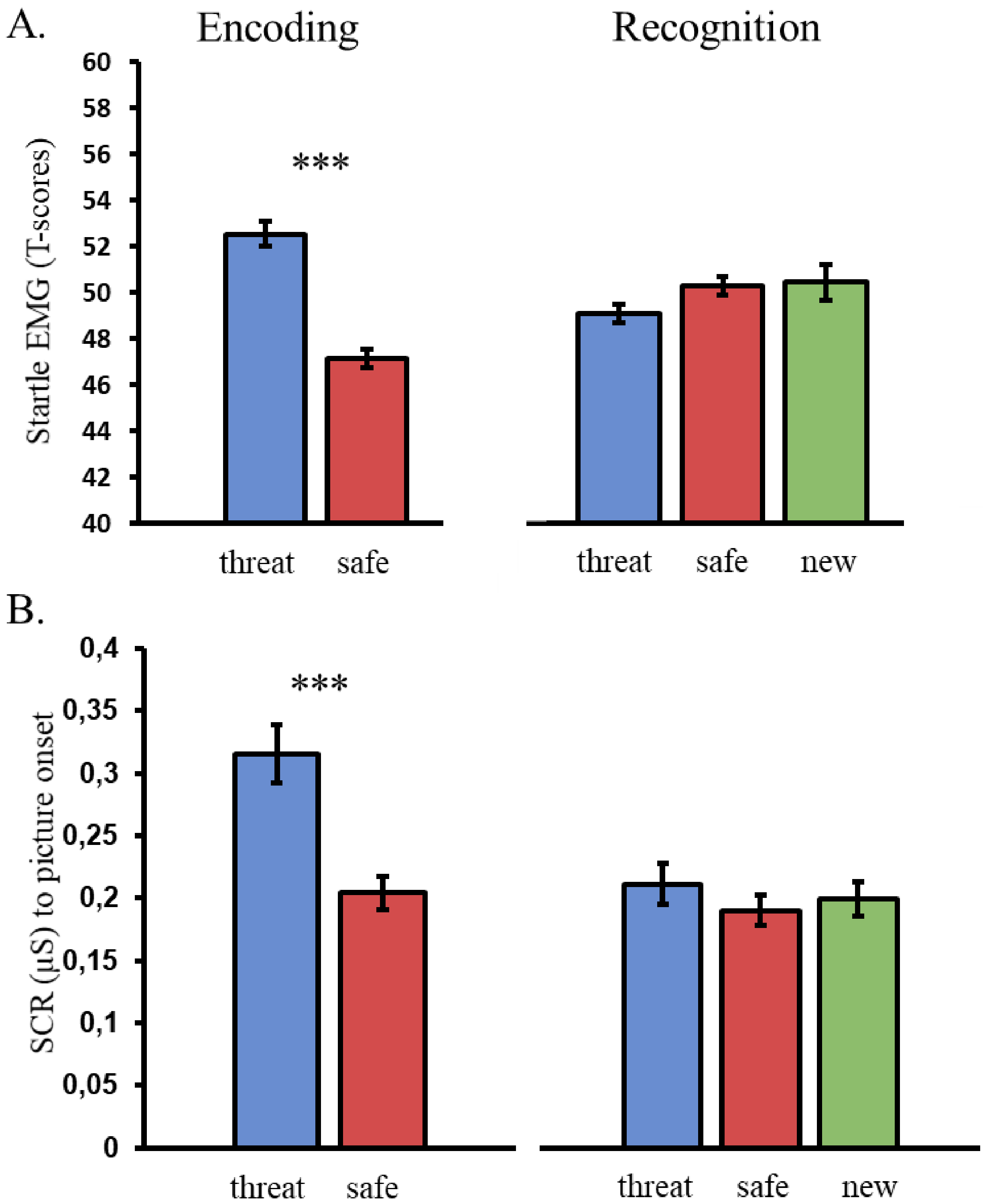

Figure 4

(A) Mean startle reflex and (B) skin conductance response to picture onset for the encoding and recognition session, as a function of contextual threat and safety, or new faces (M and SEM, *** $p < .001$).