Received: 31 August 2022 | Revised: 4 January 2023 | Accepted: 2 February 2023

DOI: 10.1111/psyp.14273

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

A partner's smile is not per se a safety signal: Psychophysiological response patterns to instructed threat and safety

Cristina Morato[1] | **Pedro Guerra**[1] | **Florian Bublatzky**[2]

[1]Department of Personality, Assessment, and Psychological Treatment, Faculty of Psychology, University of Granada, Granada, Spain

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

Correspondence
Pedro Guerra, Department of Personality, Assessment, and Psychological Treatment, Faculty of Psychology, University of Granada, Granada, Spain.
Email: pguerra@ugr.es

Florian Bublatzky, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, J5, 68159 Mannheim, Germany.
Email: florian.bublatzky@zi-mannheim.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: BU 3255/1-2; Ministerio de Ciencia e Innovación, Grant/Award Number: PID2020-119549GB-I00

Abstract
Recent studies on fear conditioning and pain perception suggest that pictures of loved ones (e.g., a romantic partner) may serve as a prepared safety cue that is less likely to signal aversive events. Challenging this view, we examined whether pictures of smiling or angry loved ones are better safety or threat cues. To this end, 47 healthy participants were verbally instructed that specific facial expressions (e.g., happy faces) cue threat of electric shocks and others cue safety (e.g., angry faces). When facial images served as threat cues, they elicited distinct psychophysiological defensive responses (e.g., increased threat ratings, startle reflex, and skin conductance responses) compared to viewing safety cues. Interestingly, instructed threat effects occurred regardless of the person who cued shock threat (partner vs. unknown) and their facial expression (happy vs. angry). Taken together, these results demonstrate the flexible nature of facial information (i.e., facial expression and facial identity) to be easily learned as signals for threat or safety, even when showing loved ones.

KEYWORDS
aversive learning, facial expression, romantic partner, startle reflex, threat-of-shock

1 | INTRODUCTION

Recent research has started to examine the beneficial effects of seeing loved ones (e.g., romantic partners, attachment figures; Coan et al., 2006; Eisenberger et al., 2011). For instance, looking at pictures of loved ones triggers a psychophysiological response pattern that is characteristic of positive emotions (e.g., pleasure ratings, increased heart rate and zygomaticus activity, as well as reduced startle reflex; Guerra, Sánchez-Adam, et al., 2012; Vico et al., 2010). Moreover, the real or pictorial presence of loved ones can mitigate the experience of pain and activate neural systems involved in reward processing (e.g., Eisenberger et al., 2011; Montoya et al., 2004; Younger et al., 2010). Regarding learning processes, however, results are inconsistent for different learning paradigms and types of dependent variables. Some studies suggest that loved ones may act as prepared safety stimuli that are less easily learned as threat cues; other studies did not find such an effect (e.g., Hornstein et al., 2016; Morato et al., 2021). Further research is needed to better understand the mechanisms involved and the circumstances under which looking at loved ones has a positive effect on learning processes.

As an important source of social support, satisfying romantic relationships have a great impact on health (e.g., lower morbidity and mortality; Holt-Lunstad et al., 2008), and even the mere pictorial presence of supportive others is helpful. For instance, reduced pain perception has been observed when having visual contact with one's loved ones (Duschek et al., 2020) or when viewing pictures of significant others (e.g., romantic partner; Eisenberger et al., 2011; Master et al., 2009). Moreover, Younger et al. (2010) examined pain perception while either viewing pictures of the romantic partner, or when completing a distraction task. Whereas the partner and the distraction task significantly reduced pain perception, only viewing the romantic partner also activated neural structures involved in reward processing (e.g., nucleus accumbens, dorsolateral prefrontal cortex; Younger et al., 2010). In addition, Eisenberger et al. (2011) found increased ventromedial prefrontal cortex activity, a key structure for processing of safety signals (for reviews, see Fullana et al., 2015; Laing et al., 2022), when viewing supportive individuals, which was also associated with a significant reduction in pain ratings and pain-related neural activity.

Fewer studies have examined the impact of loved ones on learning processes, and in most cases, classical (Pavlovian) threat conditioning has been used (e.g., Hornstein et al., 2016; Toumbelekis et al., 2018). In this procedure, a (neutral) stimulus acquires aversive qualities (conditioned stimulus, CS) through pairings with an unconditioned stimulus (US, e.g., electric shocks). Here, Seligman's preparedness theory states that some stimuli are more readily learned as threat cues, because they have evolutionary threatened humans (e.g., snakes and spiders; Hugdahl, 1978; Seligman, 1971). With regard to facial stimuli, this notion was supported by a recent meta-analysis showing small-to-moderate effects that threatening facial expressions (fear and anger) are the more potent conditioned threat cues (Ney et al., 2022). Transferring this theory to positive stimuli, some recent studies have reported that pictures of supportive others could prevent the acquisition of aversive features when used as conditioned stimuli in a classical conditioning protocol (Hornstein & Eisenberger, 2017). Thus, social support figures could act as prepared safety cues, and their targeted use as fear suppressors in research and practice could help to better understand disorders and improve clinical treatments (Hornstein et al., 2022).

While threat and safety learning and its relevance to fear and anxiety disorders are frequently examined using Pavlovian conditioning procedures (e.g., Bouton, 2002; Duits et al., 2015; Jovanovic et al., 2012; Laing & Harrison, 2021), however, much affective learning does not rely on direct first-hand experiences of aversive events (Rachman, 1977). Often, threat and safety information is socially communicated by observing the actions and experiences of others or through verbal instructions (e.g., Bandura & Walters, 1977; Olsson & Phelps, 2007). Such social learning processes provide important, but rarely studied, information on the emergence, maintenance, and treatment of (anxiety) psychopathology (Espinosa et al., 2020; Muris & Field, 2010; Schellhaas et al., 2022).

Using the instructed threat paradigm (also called threat-of-shock; Grillon et al., 1991), participants are verbally instructed that certain experimental stimuli, situations, or contextual settings signal the possibility of receiving unpleasant electric shocks (i.e., threat cues), while other stimuli serve as instructed safety cues. Recent research consistently observed that instructed threat relative to safety cues triggers aversive apprehensions, activates a neural fear network, guides selective attention, and triggers defensive physiological response systems (e.g., potentiated startle reflex, enhanced skin conductance responses [SCR]; Bublatzky & Schupp, 2012; Costa et al., 2015; Kavcıoğlu et al., 2021; Mechias et al., 2010; Mertens et al., 2016). With regard to the impact of loved familiar faces on instructed threat learning, two recent studies argue against the notion of preparedness and showed that loved face pictures are neither per se (by themselves) safe nor do reduce the effects of contextual threat (Bublatzky et al., 2022; Morato et al., 2021). However, these studies used neutral facial expressions and therefore did not consider the expressive function of a face, which is crucial for social communication of emotional states (e.g., an angry expression signals threat, a smile may indicate safety; Adolphs, 2009; Öhman et al., 2000). Thus, facial expressions are important for non-verbal communication and can contradict or even replace the spoken word (Phutela, 2015). Although there has been research on the processing of loved faces and on the processing of facial expressions, the modulatory effect of the facial expression of loved people on threat learning is poorly understood.

The present study examined whether a romantic partners´ facial expression signals greater threat or safety (e.g., an angry or smiling partner) than an unknown person's facial expression. For this purpose, pictures of beloved and unknown faces were shown expressing joy or anger, while in two experimental groups, facial expressions served as verbally instructed threat or safety cues (i.e., angry-threat group vs. happy-threat group). As in previous research, we hypothesized that loved faces are perceived as more pleasant, arousing, and less threatening. In terms of physiological responses, this will be associated with enhanced SCRs (reflecting greater arousal), increased heart rate and zygomaticus activity (two specific indices of positive emotion) for loved relative to unknown faces (Guerra et al., 2011; Guerra, Sánchez-Adam, et al., 2012; Guerra, Vico, et al., 2012; Vico et al., 2010). For the defensive startle reflex, one previous study reported that viewing loved ones may inhibit reflex activity (Guerra, Sánchez-Adam, et al., 2012), whereas other studies have failed to show this effect (Bublatzky et al., 2022; Morato et al., 2021).

While the impact of facial expression on physiological responding is mixed (e.g., Anokhin & Golosheykin, 2010; Bublatzky & Alpers, 2017), pronounced main effects of instructed threat linked to facial expressions were predicted. Specifically, viewing instructed threat cues will be associated with threat-potentiated startle reflex, enhanced SCRs, and a pronounced cardiac deceleration regardless of which facial expression cued threat (cf. Bradley et al., 2001, 2005; Bublatzky et al., 2018, 2019). Interaction effects related to which facial expression and face identity indicate threat or safety may support the notion of prepared threat and/or safety learning. For example, an angry unknown face that signals threat might show the most pronounced threat effects (prepared threat cue), or conversely, a partner's smile might be less easily learned as a threat cue (prepared safety cue).

2 | OPEN PRACTICES STATEMENT

The experiment was not preregistered, but was elaborated in a (non-public) grant application to the German Research Foundation. The data and the data-analysis scripts are posted at https://osf.io/2r8af/?view_only=6f9fa30e355c4f65a90a870e6f78be82

3 | METHOD

3.1 | Participants

Forty-seven students (36 female, mean age = 21.14 years, *SD* = 2.7) were recruited from the University of Granada (Spain). Sample size was determined based on our previous research using loved faces and instructed threat manipulations (e.g., Bublatzky et al., 2014; Guerra et al., 2011; Guerra, Sánchez-Adam, et al., 2012; Guerra, Vico, et al., 2012; Morato et al., 2021). In addition, statistical estimations with G*Power (Faul et al., 2009) indicated that a sample size of $N = 46$ was required to detect Group by face Identity effects at a medium effect size ($f = .2$, power = .90, α error = .05, and assumed correlation of repeated measures in repeated-measure ANOVAs = .5).

All participants were in good health and had normal or corrected-to-normal vision. They provided informed consent prior to their participation and received course credits. The ethics committee of the University of Granada approved the experimental protocol, which was in accordance with the Declaration of Helsinki. For some participants, individual psychophysiological data sets were lost due to technical problems (e.g., electrode failure, excessive noise). Specifically, each one data set was lost for startle blink and zygomaticus, three for heart rate, two for skin conductance and corrugator measurements, as well as one participant forgot to complete the valence rating.

3.2 | Materials and design

Facial photographs of the participants´ own romantic partner and an unknown person (another participant's partner) showing happy, neutral, and angry facial expressions were used (i.e., 6 pictures in total). All pictures were matched for size (886 × 886 pixels), color (black and white), and background (light-colored). Participants were randomly assigned to one out of two experimental groups where either happy or angry facial expressions served as threat or safety cues, whereas neutral faces remained uninstructed.

Thus, a 2 × (2 × 3) design was used, with Group (happy faces cueing threat vs. angry faces cueing threat) as a between-subject factor, and Face Identity (romantic partner vs. unknown) and Expression (happy, neutral, and angry) as repeated-measures variables. Threat and safety contingencies were verbally instructed and counterbalanced across participants. The sequence was pseudorandom with the restriction that the same face identity or emotional expression (e.g., romantic partner or happy facial expression) could not appear in more than four consecutive trials. No more than four picture-startle or picture-alone trials in a row were allowed. Importantly, and in order to focus on the impact of aversive anticipation, no shocks were delivered during the experiment. Nevertheless, a work-up procedure was carried out before the experiment started to enhance credibility of the threat-of-shock instructions (e.g., Bublatzky et al., 2019).

Psychophysiological recordings started with a 2-min rest period. Following, the experiment began with the presentation of in total 78 picture trials (each picture 13 times). Each trial consisted of a 4-s baseline period, a 6-s picture presentation, a 4-s post-picture period, and a varying inter-trial interval ranging from 2 to 4 s (see Figure 1). Auditory startle probes were equally distributed across pictures and delivered at either 4,

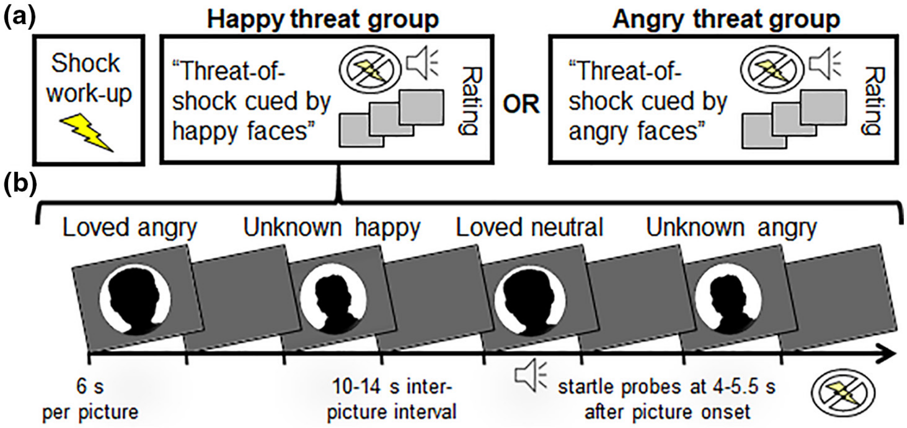


FIGURE 1 Schematic illustration of the experimental procedure. (a) A shock work-up procedure was carried out to ensure believability of the threat instructions. Participants were assigned to one of two experimental groups in which happy or angry facial expressions were instructed as threat-of-shock cues. Thus, the loved or unknown face identity was not diagnostic for shock threat. (b) For each group, the romantic partner and the control face (i.e., unknown romantic partner of another participant) were presented 39 times (78 trials in total) displaying happy, neutral, and angry facial expressions. Auditory startle probes were presented in 48 trials. No shocks were applied throughout the experiment.

4.5, 5, or 5.5 s after picture onset in 48 out of 78 trials. Additionally, eight startle probes were presented during the inter-trial intervals.

Startle probes (white noise bursts at 105 dBs, 50 ms duration and virtually instantaneous rise time) were elicited by a Coulbourn S81-02 noise generator, gated by a Coulbourn S82-24 audio-mixer amplifier (Coulbourn Instruments, Whitehall, PA) and presented over matched Telephonics TDH-49P earphones. Stimulus control was accomplished by use of Presentation software (Neurobehavioral Systems, Inc., Albany, CA), and collection of physiological data was controlled by VPM (Cook, 2001). Finally, a Letica-shock-module LI 2700 (Letica, Barcelona, Spain) was used to administer electrical pulses during the work-up procedure.

3.3 | Procedure

All participants completed an initial telephone interview to ensure that the following inclusion criteria were met: (1) having a romantic relationship, and (2) reporting a good relationship with their partner (i.e., at least 60% satisfaction on a scale from 0 to 100%). Then, specific instructions were given for taking the pictures: frontal view of the face with the three emotional expressions (happy, neutral, and angry), light background without objects, and the pictures were to be taken by someone other than the participant. Upon their arrival at the laboratory, participants filled out questionnaires on positive–negative affect (PANAS; Watson et al., 1988) and social support (MOS; Sherbourne & Stewart, 1991).

Subsequently, participants were seated in a dimly lit, sound-attenuated room, sensors were attached, and a shock work-up procedure was carried out (Bublatzky et al., 2010, 2018, 2020). This procedure included the presentation of several electrical pulses – progressively increasing in intensity – until the participants reported them as maximally unpleasant but not painful. Then, the main instructions regarding threat and safety contingencies were given, stating which emotional facial expression served as threat or safety cue (i.e., the face identity was not relevant to the threat/safety instructions). For example, in the angry-threat group, participants were told, "if you see either of these two pictures [pictures of the angry loved and unknown person were shown], there is always the possibility of receiving an electric shock as long as the picture is present." For the safety instruction, "if you see either of these two pictures [pictures of the happy loved and unknown person were shown], you will not receive any electric shock." The same instruction was given for the happy-threat group, but with the respective different facial expressions. For both groups, neutral facial expressions remained uninstructed, serving as an implicit safety cue. The pictures were presented on a 19″ flat monitor located at approximately 60 cm from the eyes of the participant. Participants' task was to watch all visual stimuli for the entire time they were on the screen.

By the end of the experiment, participants completed picture ratings. Valence and arousal ratings were assessed using the Self-Assessment Manikin (SAM; Bradley & Lang, 1994), a well-validated pictorial scale that ranges from 1 to 9 indicating unpleasant–pleasant and calm–highly aroused. The accompanying instructions were to rate "how unpleasant/pleasant (calm/aroused) this picture makes you feel." For threat ratings, participants were asked to indicate "how threatening did you find each of the faces on a scale ranging from 0 (not at all) to 10 (very threatening)?" In addition, questionnaires on empathy (Interpersonal Reactivity Index, IRI; Pérez-Albéniz et al., 2003) and attachment style (Experience of Close Relationship, ECR; Alonso-Arbiol et al., 2007) were filled out. Finally, a debriefing was held in which we asked participants a standard question, "On a scale of 0 to 10, how credible did you find the threat instruction?" Overall, participants rated the threat instruction as very credible with a mean score of 9.17 out of 10 ($SD = 1.77$), supporting the validity of threat-of-shock instructions.

3.4 | Data recording and reduction: Peripheral measures

Heart rate was derived from the electrocardiogram recorded with a Coulbourn V75-04 bio-amplifier at lead II using three standard In Vivo Metrics Ag/AgCl electrodes filled with conductive gel (Parker Laboratories, Inc, New Jersey, USA). Frequencies below 1.5 Hz and above 20 Hz were filtered out using a Coulbourn V75-48 band-pass filter. The signal was amplified by 5000 and sampled at 1000 Hz. For skin conductance, two standard Ag/AgCl electrodes were placed on the hypothenar eminence of the left hand and filled with isotonic gel (Biopac Systems). The signal was acquired with a Coulbourn V71-23 module and sampled at 50 Hz. All EMG activity was recorded by means of miniature In Vivo Metrics electrodes filled with gel and separate Coulbourn V75-04 bio-amplifiers. The raw signals were band-passed filtered (28–500 Hz), and subsequently rectified and integrated using a Coulbourn V75-24 multi-channel integrator. Raw zygomaticus and corrugator signals were sampled at 100 Hz and integrated using a 500 ms time constant. For orbicularis muscle activity, these values were set at 1000 Hz and 20 ms, respectively.

To account for temporal changes, heart rate, skin conductance, zygomaticus, and corrugator responses were calculated by averaging across each half-second for the duration of the picture display and by subtracting the activity within 1 s prior to the picture onset (e.g., Bradley et al., 2001, 2018; Costa et al., 2015; Guerra et al., 2011).

Startle responses were scored with an automated detection algorithm (Balaban et al., 1986), verified by visual inspection. The startle amplitude was defined as the difference between the peak and the onset of the response, in a time window between 20 and 120 ms after stimulus onset. To control for between-subject variability, startle amplitudes for each participant were transformed to T scores.

3.5 | Data analysis

3.5.1 | Self-report data

Valence and arousal ratings of the face pictures were analyzed by means of repeated-measures ANOVAs including the between factor Group (happy-threat vs. angry-threat), and the within factors Face Identity (romantic partner vs. unknown) and Expression (happy, neutral, angry).

3.5.2 | Peripheral measures

The effect of verbal threat/safety instructions on psychophysiological measures was assessed using repeated-measures ANOVAs with Face Identity (romantic partner vs. unknown), and Expression (happy, neutral, angry) as within-subject variables, and Group (happy-threat vs. angry-threat) as a between-subject factor. For skin conductance, heart rate, zygomaticus, and corrugator EMG, an additional factor Time (12 half-second bins) was included to examine the temporal unfolding of peripheral responses.

For all analyses, statistical significance level was set at $\alpha = 0.05$, partial eta square (η_p^2) was computed as a measure of effect size, and 95% confidence intervals are reported. Finally, Greenhouse–Geisser correction was applied to adjust for lack of sphericity in the data, and Bonferroni correction was used for pairwise comparisons.

4 | RESULTS

4.1 | Self-report data

4.1.1 | Valence

Confirming previous findings, pictures of the romantic partner were rated as significantly more pleasant than photographs depicting unknown individuals, Identity $F(1,44) = 67.89$, $p < .001$, $\eta_p^2 = .61$ (see Table 1). An interaction Identity × Group, $F(1,44) = 10.27$, $p < .01$, $\eta_p^2 = .19$, revealed that pictures of unknown individuals were rated as equally pleasant regardless of the experimental group, $p = .71$, whereas pictures of the romantic partner were significantly more pleasant for the happy-threat group, compared to the same kind of pictures in the angry-threat group, $p < .01$.

On the other hand, when considering the emotional expression being displayed, a significant linear trend was found (happy > neutral > angry), Expression $F(2,88) = 76.80$, $p < .001$, $\eta_p^2 = .64$, with significant pair-

TABLE 1 Ratings of picture valence, arousal, and perceived threat as a function of Face Identity (romantic partner vs. unknown), Group (happy-threat, angry-threat), and Facial Expression (happy, angry, and neutral).

			Valence			Arousal			Threat ratings		
Identity	**Group**	**Expression**	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Romantic partner	Happy-Threat	Happy	8.37	1.01	[7.95, 8.80]	5.62	2.39	[4.61, 6.63]	3.04	3.48	[1.54, 4.55]
		Neutral	6.46	1.84	[5.68, 7.24]	3.96	2.48	[2.91, 5.00]	1.78	2.83	[0.56, 3.00]
		Angry	5.00	2.26	[4.04, 5.96]	5.29	1.63	[4.60, 5.98]	4.30	3.20	[2.92, 5.69]
	Angry-threat	Happy	8.64	0.58	[8.38, 8.89]	5.56	2.79	[4.36, 6.77]	0.65	1.61	[−0.05, 1.35]
		Neutral	4.77	1.95	[3.91, 5.64]	4.09	1.97	[3.23, 4.94]	2.48	2.91	[1.22, 3.73]
		Angry	3.32	1.81	[2.51, 4.12]	5.78	1.70	[5.05, 6.52]	5.96	2.46	[4.89, 7.02]
Unknown	Happy-Threat	Happy	5.58	1.38	[5.00, 6.17]	4.75	2.13	[3.85, 5.65]	3.56	3.78	[1.93, 5.20]
		Neutral	4.75	0.85	[4.39, 5.11]	3.75	1.48	[3.12, 4.38]	1.91	2.21	[0.96, 2.87]
		Angry	4.33	1.83	[3.56, 5.11]	4.92	1.53	[4.27, 5.56]	4.83	3.21	[3.44, 6.22]
	Angry-threat	Happy	6.23	1.02	[5.77, 6.68]	3.78	1.70	[3.05, 4.52]	1.35	2.17	[0.41, 2.28]
		Neutral	4.32	1.13	[3.82, 4.82]	3.39	2.19	[2.44, 4.34]	3.91	3.19	[2.53, 5.29]
		Angry	3.91	1.27	[3.35, 4.47]	5.74	2.24	[4.77, 6.71]	7.17	2.44	[6.12, 8.23]

wise comparisons, all *ps* < .001. This effect varied as a function of Identity × Expression, $F(2,88) = 26.72$, $p < .001$, $\eta_p^2 = .38$, indicating that happy and neutral expressions were rated more pleasant when displayed by a loved compared to an unknown person, all *ps* < .001, but no differences for angry expressions, $p = .89$. Moreover, an interaction Expression × Group emerged, $F(2,88) = 5.90$, $p < .01$, $\eta_p^2 = .12$. For the happy-threat group, neutral and angry expressions were rated significantly more pleasant compared to the same expressions in the angry-threat group, all *ps* < .05. However, valence ratings for happy expressions did not differ between groups, $p = .07$.

4.1.2 | Arousal

Pictures of the romantic partner were rated as more arousing than unknown faces, Identity $F(1,45) = 6.49$, $p < .05$, $\eta_p^2 = .13$. However, no interaction effects were found with Group, $Fs < 1.88$, $p > .16$, $\eta_p^2 < .04$.

With regard to emotional facial expressions, a main effect of Expression was observed, $F(2,90) = 14.95$, $p < .001$, $\eta_p^2 = .25$, with happy and angry expression rated as more arousing than neutral faces, all *ps* < .001. Moreover, an interaction Identity × Expression emerged, $F(2,90) = 3.44$, $p < .05$, $\eta_p^2 = .07$, showing that for partner pictures, happy and angry expressions were rated as more arousing compared to neutral, all *ps* < .001. In contrast, unknown faces displaying angry expressions were more arousing than both neutral and happy expressions, all *ps* < .05. No differences were found between happy and neutral expressions in this subset of photographs, $p = .19$.

4.1.3 | Threat ratings

Pictures of unknown individuals were rated as more threatening compared to photographs of the romantic partner, Identity $F(1,44) = 16.00$, $p < .001$, $\eta_p^2 = .27$. The interaction Identity × Group did not reach significance, $F(1,44) = 3.7$, $p = .06$, $\eta_p^2 = .08$.

Moreover, a main effect of Expression emerged, $F(2,88) = 23.25$, $p < .001$, $\eta_p^2 = .35$, showing angry expressions as more threatening than both neutral and happy faces, all *ps* < .001. This effect varied as a function of the experimental group, Expression × Group $F(2,88) = 8.92$, $p < .001$, $\eta_p^2 = .17$. In the happy-threat group, only angry expressions were rated significantly more threatening than neutral faces, $p < .001$, but no difference emerged for the other comparisons, *ps* > .43. For the angry-threat group, however, both neutral and angry faces prompted significantly larger threat ratings as compared to happy expressions, all $ps < .05$. Finally, angry expressions were also significantly more threatening than neutral faces, $p < .001$.

4.2 | Startle reflex

Contrary to our expectation, neither a main effect of Face Identity (partner vs. unknown), $F(1,44) = .11$, $p = .75$, $\eta_p^2 < .01$, nor an interaction Identity × Group, $F(1,44) = .26$, $p = .62$, $\eta_p^2 = .01$, was found for the startle reflex.

Moreover, a main effect of Expression was found, $F(2,88) = 8.7$, $p < .001$, $\eta_p^2 = .16$, indicating potentiated startle reflex when viewing angry and happy facial expressions relative to neutral expressions, all $ps < .01$, but no difference between angry and happy faces, $p = 1.0$. Interestingly, this effect varied with threat instructions as evidenced by a significant Expression × Group interaction, $F(2,88) = 18.18$, $p < .001$, $\eta_p^2 = .29$. As shown in Figure 2a (left panel), happy faces cueing threat potentiated the startle reflex compared to both neutral (uninstructed) and angry faces (cueing safety), all $ps < .01$. Similarly, angry faces cueing threat (Figure 2a, right panel) prompted significantly larger startle responses compared to happy (safety cues) and neutral (uninstructed) expressions, all $ps < .001$. For both experimental groups, no differences were found between neutral and instructed safety expressions, all $ps > .74$.

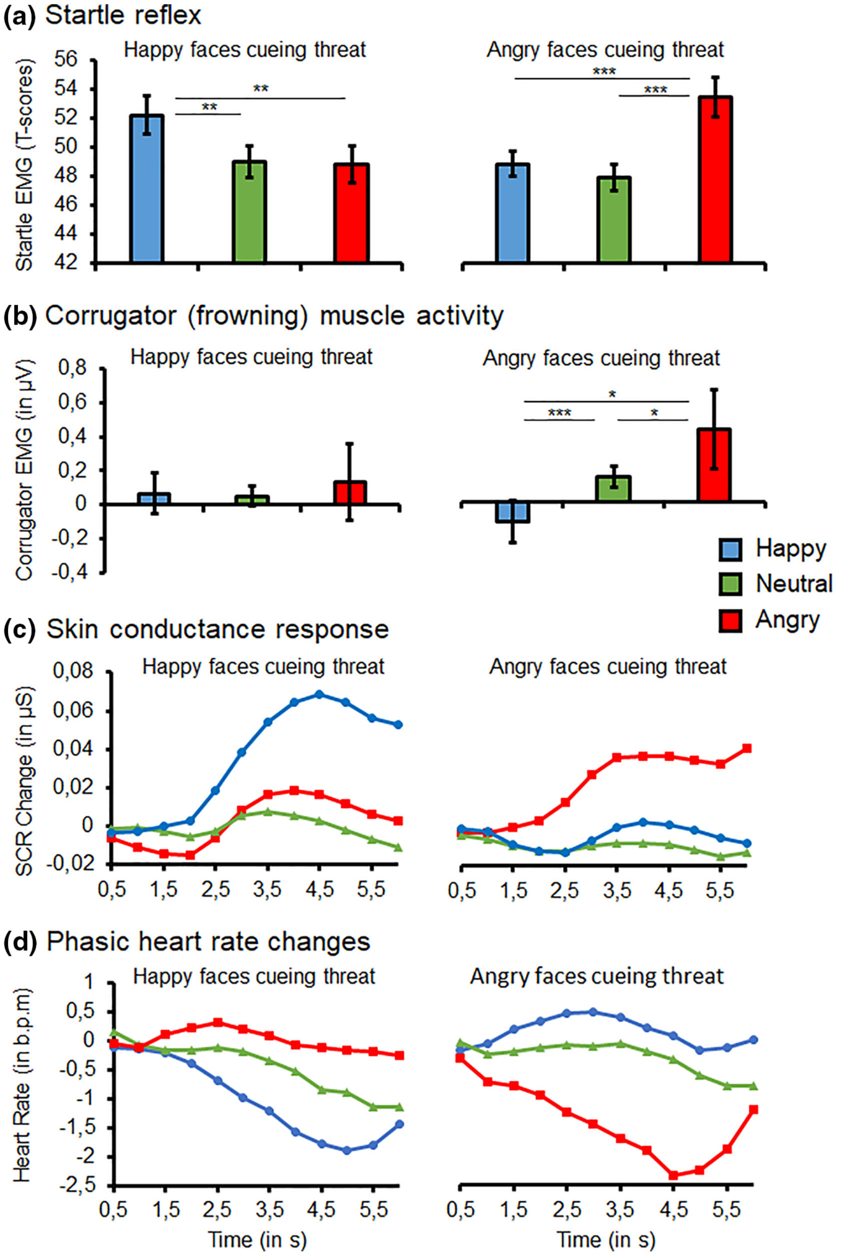


FIGURE 2 Mean amplitudes (95% CI) of the (a) startle reflex and (b) corrugator EMG show significant interaction effects as a function of facial Expression × Group (happy-threat vs. angry-threat signals). (c) Skin conductance responses and (d) phasic heart rate changes reveal this interaction effect across time (Expression × Group × Time). $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

4.3 | Corrugator EMG

An increase of corrugator (frowning) activity was observed for unknown relative to loved faces, Identity $F(1,43) = 19.65$, $p < .001$, $\eta_p^2 = .31$. This effect developed across time, Identity × Time $F(11,473) = 5.72$, $p < .001$, $\eta_p^2 = .12$, and showed significant differences at all time points, all *ps* < .01. There was no interaction of Identity × Group, $F(1,43) = .45$, $p = .51$, $\eta_p^2 = .01$.

As shown in Figure 2b, corrugator activity varied as a joint function of emotional expression and threat instruction, Expression × Group $F(2,86) = 4.47$, $p < .05$, $\eta_p^2 = .09$. Separate analysis of the angry-threat group showed increased corrugator activity for angry relative to neutral and happy expressions, all *ps* < .05, whereas happy expressions (signaling safety) were inhibited corrugator activity relative to neutral pictures, $p < .001$. In the happy-threat group, corrugator activity did not vary between facial expressions.

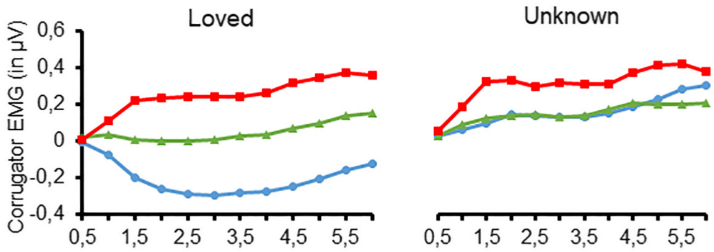
Moreover, an interaction Identity × Expression × Time emerged, $F(22,946) = 4.56$, $p < .01$, $\eta_p^2 = .10$ (see Figure 3a). Angry expressions of the partner elicited increased corrugator activity, whereas smiling partner pictures inhibited the frowning response. Pairwise comparisons revealed significant differences between happy and angry expressions (seconds 1 to 5.5), happy and neutral (seconds 1 to 6), and between angry and neutral faces (seconds 1.5 to 2), all *ps* < .05. In contrast, pictures of unknown faces elicited increased corrugator activity that was particularly pronounced for angry relative to neutral and happy expressions. Pairwise comparisons for each time point separately showed significantly enhanced activity for angry relative to happy (seconds 1 to 1.5) and neutral faces (seconds 1.5 to 3.5 and 5 to 5.5), all *ps* < .05; no differences were found between happy and neutral expressions, all *ps* > .28.

4.4 | Zygomaticus EMG

The zygomaticus (smiling) activity was enhanced when viewing pictures of the partner as compared to unknown faces (see Figure 3b), Identity $F(1,44) = 19.91$ $p = .001$, $\eta_p^2 = .31$. This difference developed across time, Identity × Time $F(11,484) = 17.57$, $p < .001$, $\eta_p^2 = .29$, with significant differences from 1 to 6 s after picture onset, all *ps* < .05. Interestingly, there were no main or interaction effects with Group, $Fs < 2.60$, $p > .10$, $\eta_p^2 < .06$, suggesting that zygomaticus muscle was the only measure not modulated by threat instructions.

With regard to emotional facial expressions, the zygomaticus activity revealed a main effect of Expression, $F(2,88) = 10.93$, $p < .001$, $\eta_p^2 = .20$, and significant interactions Expression × Time, $F(22,968) = 7.05$, $p < .001$, $\eta_p^2 = .14$, Identity × Expression, $F(2,88) = 8.64$, $p < .01$, $\eta_p^2 = .16$, as well as Identity × Expression × Time, $F(22,968) = 6.13$ $p < .01$, $\eta_p^2 = .12$. Following-up the latter three-way interaction, pairwise comparisons show that partner pictures displaying a smile prompted significantly larger responses relative to neutral (secs 1 to 6) and angry faces (secs 1.5 to 6), all *ps* < .05. Angry and neutral partner expressions differed significantly from seconds 1 to 1.5, all *ps* < .05. For pictures of unknown faces, larger zygomaticus activity was observed for happy compared to neutral faces from seconds 2 to 4, all *ps* < .05.

(a) Corrugator (frowning) muscle activity



(b) Zygomaticus (smiling) muscle activity

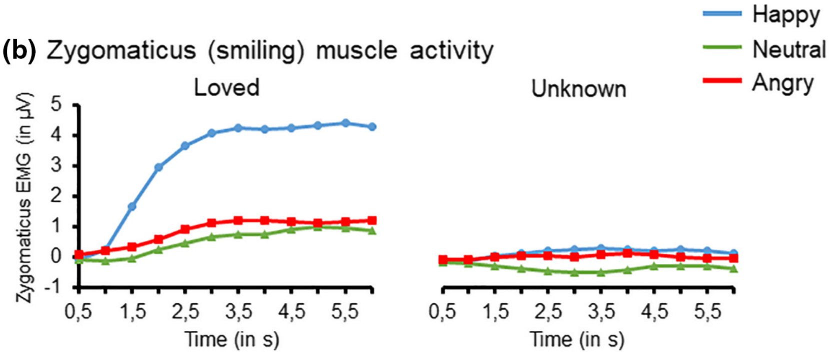


FIGURE 3 Mean amplitudes of the (a) corrugator and (b) zygomaticus muscle activity as a function of Identity × Expression × Time.

4.5 | Skin conductance

For the skin conductance response, neither the main effect Face Identity, $F(1,43) = .28$, $p = .60$, $\eta_p^2 = .01$, nor any interaction involving Identity was significant, $Fs < 1.75$, $ps > .19$, $\eta_p^2 < .04$.

In terms of emotional facial expressions, an interaction Expression × Group was observed, $F(2,86) = 4.07$, $p < .05$, $\eta_p^2 = .09$, which further tended to vary across time (see Figure 2c), Expression × Group × Time $F(22,946) = 3.18$, $p = .05$, $\eta_p^2 = .07$. For the happy-threat group, enhanced SCRs were observed for happy versus neutral faces (seconds 4.5 to 6), all $ps < .05$, but no difference for the other comparisons, all $ps > .054$. With regard to the angry-threat group, SCRs were enhanced for angry compared to neutral expressions (seconds 4 to 5), all $ps < .05$, no significant differences emerged for the other time points and comparisons, all $ps > .052$.

4.6 | Heart rate

Phasic heart rate changes did not vary with Face Identity or related interactions, $Fs < 2.22$, $ps > .12$, $\eta_p^2 < .05$. As illustrated in Figure 2d, pronounced HR changes emerged as a joint function of Expression × Time × Group, $F(22,924) = 4.67$, $p < .001$, $\eta_p^2 = .1$. In the happy-threat group, pairwise comparisons for each time point show significant HR deceleration when viewing happy compared to angry faces (from 4 to 6 s after picture onset), all $ps < .05$. No further comparison reached significance, all $ps > .08$.

For the angry-threat group, a pronounced HR deceleration was observed when viewing angry faces compared to happy (1.5 to 6 s) and neutral expressions (2.5 to 5 s), all $ps < .05$. No differences were found in this group between happy and neutral faces, all $ps > .33$.

5 | DISCUSSION

The present study examined whether happy and angry facial expressions of a romantic partner or an unknown person differentially serve as threat or safety cues. Results show that threat learning through verbal instructions leads to pronounced psychophysiological defensive responses. Specifically, instructed threat faces were rated as more unpleasant, arousing, and threatening relative to safety cues. Regarding physiological response markers, threat-potentiated startle reflex and more corrugator frowning activity was found for threat relative to safety cues. Moreover, enhanced skin conductance responses and pronounced heart rate deceleration were observed for instructed threat cues. Importantly, all threat effects occurred regardless of face identity and expression, suggesting that happy and angry facial expressions of loved and unknown persons served equally well as instructed threat cues. Together, these data indicate that pictures of (smiling) loved ones do neither constitute a prepared safety signal nor impede threat learning.

When told that someone might be dangerous (i.e., associated with aversive events), psychophysiological defense systems are activated when confronted with that person. The present findings support this notion and replicate previous work using facial stimuli and the instructed threat-of-shock paradigm (Bublatzky et al., 2018, 2019, 2020; Grillon & Charney, 2011; Morato et al., 2021). Specifically, threat cues effectively activated the somatic (i.e., potentiated startle reflex and increased corrugator “frowning” activity) and the autonomic nervous system (i.e., enhanced skin conductance responses and phasic deceleration of the heart rate). Importantly, and replicating one previous study, this defense activation emerged regardless of whether the romantic partner or an unknown individual indicated threat or safety (Morato et al., 2021). Moreover, happy and angry facial expressions were equally capable of cueing instructed threat (cf. Bublatzky et al., 2018, 2019). Thus, the intrinsic affective meaning of a face picture (e.g., displaying a loved person smiling) can be readily changed by verbal instructions and trigger defensive response programs. The only physiological measure that did not vary with threat/safety instructions was the zygomatic muscle, which reflects no modulation of the smiling activity in the face of threat (Morato et al., 2021). Thus, physiological systems that are involved in defensive and withdrawal behavior faithfully respond to potentially threatening situations. Regardless of which person and/or facial expression signals threat to the observer, a “better safe than sorry” strategy appears adaptive.

Subjective ratings and physiological responding partly converged in response to instructed threat learning. For instance, when a smiling person signals threat, this person is considered as more pleasant than an angry facial expression. On the physiological side, however, the startle reflex is similarly potentiated for happy and angry facial expressions cueing threat (similar to affective scenes; Bradley et al., 2005). Such partial incongruence between subjective (self-reported) and physiological response levels is a well-known phenomenon (e.g., Lang et al., 1997; Morato et al., 2021) and may reflect controversial situations in which one thinks “all is good” but still experiences strong bodily symptoms (e.g., during a panic attack). Such discrepancies have been found also with other measures varying as a function of social support or the physical presence of others. For instance, in a cold-pressor pain task, attenuated subjective pain has been observed during a social support condition with the presence of another person (Duschek et al., 2020; Edwards et al., 2017).

Looking at pictures of loved ones has a great impact on the functioning of the somatic and autonomic nervous systems. In a safe context, viewing loved ones is consistently associated with increased zygomaticus "smiling" and decreased corrugator "frowning" activity (Guerra et al., 2011; Guerra, Vico, et al., 2012; Vico et al., 2010). One study even found attenuated defensive startle reflex when viewing loved ones (Guerra, Sánchez-Adam, et al., 2012). However, this latter effect could not be replicated in recent work as well as the present study (Bublatzky et al., 2022; Morato et al., 2021). In addition, enhanced skin conductance responses and a triphasic change in heart rate (deceleration-acceleration-deceleration) indicate changes in autonomic nervous system activity when viewing pleasant stimuli (e.g., erotic scenes, loved faces or names; Bradley et al., 2001; Guerra, Vico, et al., 2012, Guerra, Sánchez-Adam, et al., 2012; Lucas et al., 2019). Within a threatening context, however, this response pattern changes and a potentiated startle reflex, an increased corrugator activity, and a pronounced heart rate deceleration were observed, indicating defensive response preparation (Bublatzky et al., 2022; Morato et al., 2021). Here, the zygomatic muscle was the only measure that was not modulated by threat instructions, reflecting that this muscle is involved in approach-related but not threat-related facial behavior.

Consistent with these findings, our study did not support the notion that looking at one's own smiling (or angry) romantic partner reduced (or enhanced) the impact of instructed threat. In fact, when serving as threat cues, pictures of unknown faces as well as loved ones led to similarly pronounced threat effects (i.e., potentiated startle reflex, enhanced skin conductance responses, and deceleration of the heart rate; cf. Bublatzky et al., 2022; Morato et al., 2021). These findings are in contrast with recent studies that suggested attachment figures could act as prepared safety cues. For instance, Hornstein et al. (2016, 2017) reported that supportive others were more difficult to acquire aversive features using a classical conditioning protocol. While diverging findings may relate to different experimental protocols (e.g., classical conditioning vs. instructional learning; selection criteria of loved ones or supportive others), accumulating evidence suggests that facial images of loved ones are not shielded from becoming socially learned threat cues (see also Morato et al., 2021).

From a more general learning perspective, several aspects of the present study design are particularly noteworthy in addition/contrast to classical (Pavlovian) conditioning procedures. First, using the threat-of-shock paradigm, participants are explicitly informed and therefore fully aware about threat/safety contingencies. Closely related, second, is that these verbal threat/safety instructions work without shock application during the experiment, and even without shock work-up procedures (e.g., Arnold et al., 2021; Costa et al., 2015; Schellhaas et al., 2020, 2022). Thus, threat learning occurs at a 0% shock reinforcement rate, and accordingly, the omission of shocks during the course of the experiment does not necessarily lead to rapid extinction learning, as in classical conditioning designs (depending on the reinforcement schedule). In fact, instructed threat effects have been shown to persist within and even across repeated test days without experiencing any electric shocks (Bublatzky et al., 2013, 2014, 2022). Third, the cognitive representation of threat associations implies a strong anticipatory component and, accordingly, a high degree of uncertainty and prediction error regarding the occurrence of shocks (e.g., Atlas, 2019; Holland & Schiffino, 2016; Iordanova et al., 2021). Thus, the instructed threat paradigm is particularly useful for examining the effects of an uncertain or presumed threat, as seen, for example, in worry and apprehension in generalized and anticipatory anxiety, but also in stereotyping and prejudice towards strangers (Amodio, 2014; Bublatzky et al., 2020; Golkar & Olsson, 2017).

Finally, our findings on instructed threat and safety learning add to recent research on Pavlovian safety learning (for reviews see Grasser & Jovanovic, 2021; Laing & Harrison, 2021; Laing et al., 2022), pointing to the power of "mere" verbal instructions to override prior affective meaning. Here, the use of personalized stimulus materials can contribute to understanding the workings of safety learning. For example, studies have shown that fear-relevant stimuli can be conditioned as threat inhibitors even in animal-phobic participants (e.g., snakes as a safety signal; McNally & Reiss, 1982, 1984; Wilkinson et al., 1989); similarly, the present results show that safety-relevant loved faces can also be learned as threat cues. Thus, if the concept of preparedness were applicable to images of loved familiar people, verbal instructions could easily override this hypothetical preparedness. A focus on the social-affective processes involved in the extinction of cognitive aspects of fear and anxiety (e.g., social support, interaction style, relationship status and quality) may be particularly informative for improving treatment-resistant anxious and stress-related psychopathology.

Apart from learning mechanisms, several studies have reported positive effects of viewing loved ones. For instance, lower physical pain perception (Duschek et al., 2020; Master et al., 2009; Younger et al., 2010), reduced endocrine stress responses (Eisenberger et al., 2007), and better recovery from medical intervention have been observed in the presence of social supporters (e.g., Cacioppo & Cacioppo, 2014; House et al., 1988). Here, future research could explore the question of what makes a social relationship supportive. Is it the physical presence or absence, the type of prosocial or helping behavior (e.g., social touch; Eckstein et al., 2020; Goldstein et al., 2016), or the identity of the person offering support (e.g., romantic partner or parents)?

In summary, the present study examined psychophysiological defensive responses as a function of personal relevant faces, facial expression, and social threat and safety learning. Main findings reveal that neither the personal relevance of a face (displaying a loved or an unknown person) nor its facial expression serving as threat/safety cue (smile or angry expression) modulates defense activation. Specifically, the smiling romantic partner as well as the angry unknown face served equally well as threat or safety signals. Together, these findings demonstrate the flexible nature of facial information that can be readily learned as signals for threat or safety even when displaying loved ones.

AUTHOR CONTRIBUTIONS

Cristina Morato: Data curation; formal analysis; investigation; software; validation; writing – original draft; writing – review and editing. **Pedro Guerra:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; supervision; validation; writing – review and editing. **Florian Bublatzky:** Conceptualization; formal analysis; funding acquisition; methodology; project administration; resources; supervision; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGMENTS

We are grateful to Raquel Rayo for her help with data collection.

FUNDING INFORMATION

This research was supported by a grant from the German Research Foundation to FB (BU 3255/1-2) as well as by grant to PG from the Spanish Ministry of Science and Innovation (PID2020-119549GB-I00).

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY STATEMENT

The data and the data-analysis scripts are posted at https://osf.io/2r8af/?view_only=6f9fa30e355c4f65a90a870e6f78be82.

ORCID

Florian Bublatzky https://orcid.org/0000-0002-7726-0682

REFERENCES

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Alonso-Arbiol, I., Balluerka, N., & Shaver, P. R. (2007). A Spanish version of the experiences in close relationships (ECR) adult attachment questionnaire. *Personal Relationships*, *14*(1), 45–63. https://doi.org/10.1111/j.1475-6811.2006.00141.x

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670–682. https://doi.org/10.1038/nrn3800

Anokhin, A. P., & Golosheykin, S. (2010). Startle modulation by affective faces. *Biological Psychology*, *83*(1), 37–40. https://doi.org/10.1016/j.biopsycho.2009.10.001

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

Atlas, L. Y. (2019). How instructions shape aversive learning: Higher order knowledge, reversal learning, and the role of the amygdala. *Current Opinion in Behavioral Sciences*, *26*, 121–129. https://doi.org/10.1016/j.cobeha.2018.12.008

Balaban, M., Losito, B., Simons, R. F., & Graham, F. K. (1986). Off-line latency and amplitude scoring of the human reflex eyeblink with Fortran IV. *Psychophysiology*, *23*, 612.

Bandura, A., & Walters, R. H. (1977). *Social learning theory* (Vol. *1*). Prentice Hall.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–986. https://doi.org/10.1016/s0006-3223(02)01546-9

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276–298. https://doi.org/10.1037/1528-3542.1.3.276

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

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

Bradley, M. M., Moulder, B., & Lang, P. J. (2005). When good things go bad: The reflex physiology of defense. *Psychological Science*, *16*(6), 468–473.

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

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

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

Bublatzky, F., Guerra, P., & Alpers, G. W. (2018). Verbal instructions override the meaning of facial expressions. *Scientific Reports*, *8*, 14988. https://doi.org/10.1038/s41598-018-33269-2

Bublatzky, F., Guerra, P., & Alpers, G. W. (2020). Watch out, he's dangerous! Electrocortical indicators of selective visual attention to allegedly threatening persons. *Cortex*, *131*, 164–178. https://doi.org/10.1016/j.cortex.2020.07.009

Bublatzky, F., Guerra, P. M., Pastor, M. C., Schupp, H. T., & Vila, J. (2013). Additive effects of threat-of-shock and picture valence on startle reflex modulation. *PLoS One*, *8*(1), e54003. https://doi.org/10.1371/journal.pone.0054003

Bublatzky, F., Riemer, M., & Guerra, P. (2019). Reversing threat to safety: Incongruence of facial emotions and instructed threat modulates conscious perception but not physiological responding. *Frontiers in Psychology*, *10*, 2091.

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/10.21203/rs.3.rs-965841/v1

Bublatzky, F., & Schupp, H. T. (2012). Pictures cueing threat: Brain dynamics in viewing explicitly instructed danger cues. *Social Cognitive and Affective Neuroscience*, *7*(6), 611–622.

Cacioppo, J. T., & Cacioppo, S. (2014). Social relationships and health: The toxic effects of perceived social isolation. *Social and Personality Psychology Compass*, *8*(2), 58–72. https://doi.org/10.1111/spc3.12087

Coan, J. A., Schaefer, H. S., & Davidson, R. J. (2006). Lending a hand: Social regulation of the neural response to threat. *Psychological Science*, *17*(12), 1032–1039. https://doi.org/10.1111/j.1467-9280.2006.01832.x

Cook, E. W. (2001). *VPM reference manual*. Author.

Costa, V. D., Bradley, M. M., & Lang, P. J. (2015). From threat to safety: Instructed reversal of defensive reactions. *Psychophysiology*, *52*(3), 325–332. https://doi.org/10.1111/psyp.12359

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M. A., & Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety*, *32*(4), 239–253. https://doi.org/10.1002/da.22353

Duschek, S., Nassauer, L., Montoro, C. I., Bair, A., & Montoya, P. (2020). Dispositional empathy is associated with experimental pain reduction during provision of social support by romantic partners. *Scandinavian Journal of Pain*, *20*(1), 205–209. https://doi.org/10.1515/sjpain-2019-0025

Eckstein, M., Mamaev, I., Ditzen, B., & Sailer, U. (2020). Calming effects of touch in human, animal, and robotic interaction-scientific state-of-the-art and technical advances. *Frontiers in Psychiatry*, *11*, 1–17. https://doi.org/10.3389/fpsyt.2020.555058

Edwards, R., Eccleston, C., & Keogh, E. (2017). Observer influences on pain: An experimental series examining same-sex and opposite-sex friends, strangers, and romantic partners. *Pain*, *158*(5), 846–855. https://doi.org/10.1097/j.pain.0000000000000840

Eisenberger, N. I., Master, S. L., Inagaki, T. K., Taylor, S. E., Shirinyan, D., Lieberman, M. D., & Naliboff, B. D. (2011). Attachment figures activate a safety signal-related neural region and reduce pain experience. *Proceedings of the National Academy of Sciences*, *108*(28), 11721–11726. https://doi.org/10.1073/pnas.1108239108

Eisenberger, N. I., Taylor, S. E., Gable, S. L., Hilmert, C. J., & Lieberman, M. D. (2007). Neural pathways link social support to attenuated neuroendocrine stress responses. *NeuroImage*, *35*(4), 1601–1612. https://doi.org/10.1016/j.neuroimage.2007.01.038

Espinosa, L., Lundin Kleberg, J., Hofvander, B., Berggren, S., Bölte, S., & Olsson, A. (2020). Enhanced social learning of threat in adults with autism. *Molecular Autism*, *11*(1), 1–9. https://doi.org/10.1186/s13229-020-00375-w

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2015). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*, 500–508. https://doi.org/10.1038/mp.2015.88

Goldstein, P., Shamay-Tsoory, S. G., Yellinek, S., & Weissman-Fogel, I. (2016). Empathy predicts an experimental pain reduction during touch. *The Journal of Pain*, *17*(10), 1049–1057. https://doi.org/10.1016/j.jpain.2016.06.007

Golkar, A., & Olsson, A. (2017). The interplay of social group biases in social threat learning. *Scientific Reports*, *7*(1), 1–5. https://doi.org/10.1038/s41598-017-07522-z

Grasser, L. R., & Jovanovic, T. (2021). Safety learning during development: Implications for development of psychopathology. *Behavioural Brain Research*, *408*, 113297. https://doi.org/10.1016/j.bbr.2021.113297

Grillon, C., & Charney, D. R. (2011). In the face of fear: Anxiety sensitizes defensive responses to fearful faces. *Psychophysiology*, *48*(12), 1745–1752.

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

Guerra, P., Campagnoli, R. R., Vico, C., Volchan, E., Anllo-Vento, L., & Vila, J. (2011). Filial versus romantic love: Contributions from peripheral and central electrophysiology. *Biological Psychology*, *88*(2–3), 196–203. https://doi.org/10.1016/j.biopsycho.2011.08.002

Guerra, P., Vico, C., Campagnoli, R., Sánchez, A., Anllo-Vento, L., & Vila, J. (2012). Affective processing of loved familiar faces: Integrating central and peripheral electrophysiological measures. *International Journal of Psychophysiology*, *85*(1), 79–87. https://doi.org/10.1016/j.ijpsycho.2011.06.004

Guerra, P. M., Sánchez-Adam, A., Anllo-Vento, L., Ramirez, I., & Vila, J. (2012). Viewing loved faces inhibits defense reactions: A health-promotion mechanism? *PLoS One*, *7*(7), e41631. https://doi.org/10.1371/journal.pone.0041631

Holland, P. C., & Schiffino, F. L. (2016). Mini-review: Prediction errors, attention and associative learning. *Neurobiology of Learning and Memory*, *131*, 207–215. https://doi.org/10.1016/j.nlm.2016.02.014

Holt-Lunstad, J., Birmingham, W., & Jones, B. Q. (2008). Is there something unique about marriage? The relative impact of marital status, relationship quality, and network social support on ambulatory blood pressure and mental health. *Annals of Behavioral Medicine*, *35*(2), 239–244. https://doi.org/10.1007/s12160-008-9018-y

Hornstein, E. A., Craske, M. G., Fanselow, M. S., & Eisenberger, N. I. (2022). Reclassifying the unique inhibitory properties of social-support-figures: A roadmap for exploring prepared fear suppression. *Biological Psychiatry*, *91*(9), 778–785. https://doi.org/10.1016/j.biopsych.2021.11.017

Hornstein, E. A., & Eisenberger, N. I. (2017). Unpacking the buffering effect of social support figures: Social support attenuates fear acquisition. *PLoS One*, *12*(5), e0175891. https://doi.org/10.1371/journal.pone.0175891

Hornstein, E. A., Fanselow, M. S., & Eisenberger, N. I. (2016). A safe haven: Investigating social-support figures as prepared safety stimuli. *Psychological Science*, *27*(8), 1051–1060. https://doi.org/10.1177/0956797616646580

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, *241*(4865), 540–545. https://doi.org/10.1126/science.3399889

Hugdahl, K. (1978). Electrodermal conditioning to potentially phobic stimuli: Effects of instructed extinction. *Behaviour Research and Therapy*, *16*(5), 315–321. https://doi.org/10.1016/0005-7967(78)90001-3

Iordanova, M. D., Yau, J. O. Y., McDannald, M. A., & Corbit, L. H. (2021). Neural substrates of appetitive and aversive prediction error. *Neuroscience & Biobehavioral Reviews*, *123*, 337–351. https://doi.org/10.1016/j.neubiorev.2020.10.029

Jovanovic, T., Kazama, A., Bachevalier, J., & Davis, M. (2012). Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology*, *62*(2), 695–704. https://doi.org/10.1016/j.neuropharm.2011.02.023

Kavcıoğlu, F. C., Bublatzky, F., Pittig, A., & Alpers, G. W. (2021). Instructed threat enhances threat perception in faces. *Emotion*, *21*(2), 419–429. https://doi.org/10.1037/emo0000708

Laing, P. A., & Harrison, B. J. (2021). Safety learning and the Pavlovian conditioned inhibition of fear in humans: Current state and future directions. *Neuroscience & Biobehavioral Reviews*, *127*, 659–674. https://doi.org/10.1016/j.neubiorev.2021.05.014

Laing, P. A. F., Felmingham, K. L., Davey, C. G., & Harrison, B. J. (2022). The neurobiology of Pavlovian safety learning: Towards an acquisition-expression framework. *Neuroscience & Biobehavioral Reviews*, *142*, 104882. https://doi.org/10.1016/j.neubiorev.2022.104882

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97–135). Erlbaum.

Lucas, I., Sánchez-Adam, A., Vila, J., & Guerra, P. (2019). Positive emotional reactions to loved names. *Psychophysiology*, *56*(7), e13363. https://doi.org/10.1111/psyp.13363

Master, S. L., Eisenberger, N. I., Taylor, S. E., Naliboff, B. D., Shirinyan, D., & Lieberman, M. D. (2009). A picture's worth: Partner photographs reduce experimentally induced pain. *Psychological Science*, *20*(11), 1316–1318. https://doi.org/10.1111/j.1467-9280.2009.02444.x

McNally, R. J., & Reiss, S. (1982). The preparedness theory of phobias and human safety-signal conditioning. *Behaviour Research and Therapy*, *20*(2), 153–159. https://doi.org/10.1016/0005-7967(82)90114-0

McNally, R. J., & Reiss, S. (1984). The preparedness theory of phobias: The effects of initial fear level on safety-signal conditioning to fear-relevant stimuli. *Psychophysiology*, *21*(6), 647–652. https://doi.org/10.1111/j.1469-8986.1984.tb00251.x

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

Mertens, G., Raes, A. K., & De Houwer, J. (2016). Can prepared fear conditioning result from verbal instructions? *Learning and Motivation*, *53*, 7–23. https://doi.org/10.1016/j.lmot.2015.11.001

Montoya, P., Larbig, W., Braun, C., Preissl, H., & Birbaumer, N. (2004). Influence of social support and emotional context on pain processing and magnetic brain responses in fibromyalgia. *Arthritis and Rheumatism*, *50*(12), 4035–4044. https://doi.org/10.1002/art.20660

Morato, C., Guerra, P., & Bublatzky, F. (2021). Verbal threat learning does not spare loved ones. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-84921-3

Muris, P., & Field, A. P. (2010). The role of verbal threat information in the development of childhood fear."beware the Jabberwock!". *Clinical Child and Family Psychology Review*, *13*(2), 129–150. https://doi.org/10.1007/s10567-010-0064-1

Ney, L. J., O'Donohue, M. P., Lowe, B. G., & Lipp, O. V. (2022). Angry and fearful compared to happy or neutral faces as conditional stimuli in human fear conditioning: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *139*, 104756. https://doi.org/10.1016/j.neubiorev.2022.104756

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 533–575). Cambridge University Press.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102.

Pérez-Albéniz, A., De Paúl, J., Etxeberría, J., Montes, M. P., & Torres, E. (2003). Adaptación de interpersonal reactivity index (IRI) al español. *Psicothema*, *15*(2), 267–272.

Phutela, D. (2015). The importance of non-verbal communication. *IUP Journal of Soft Skills*, *9*(4), 43.

Rachman, S. (1977). The conditioning theory of fear acquisition: A critical examination. *Behaviour Research and Therapy*, *15*(5), 375–387.

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

Schellhaas, S., Schmahl, C., & Bublatzky, F. (2022). Social threat and safety learning in individuals with adverse childhood experiences: Electrocortical evidence on face processing, recognition, and working memory. *European Journal of Psychotraumatology*, *13*(2), 2135195. https://doi.org/10.1080/20008066.2022.2135195

Seligman, M. E. (1971). Phobias and preparedness. *Behavior Therapy*, *2*(3), 307–320. https://doi.org/10.1016/s0005-7894(71)80064-3

Sherbourne, C. D., & Stewart, A. L. (1991). The MOS social support survey. *Social Science & Medicine*, *32*(6), 705–714. https://doi.org/10.1016/0277-9536(91)90150-b

Toumbelekis, M., Liddell, B. J., & Bryant, R. A. (2018). Thinking of attachment figures blocks differential fear conditioning. *Social Cognitive and Affective Neuroscience*, *13*(9), 989–994. https://doi.org/10.1093/scan/nsy065

Vico, C., Guerra, P., Robles, H., Vila, J., & Anllo-Vento, L. (2010). Affective processing of loved faces: Contributions from peripheral and central electrophysiology. *Neuropsychologia*, *48*(10), 2894–2902. https://doi.org/10.1016/j.neuropsychologia.2010.05.031

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilkinson, G. M., Lovibond, P. F., Siddle, D. A., & Bond, N. (1989). Effects of fear-relevance on electrodermal safety signal learning. *Biological Psychology*, *28*(2), 89–104. https://doi.org/10.1016/0301-0511(89)90092-6

Younger, J., Aron, A., Parke, S., Chatterjee, N., & Mackey, S. (2010). Viewing pictures of a romantic partner reduces experimental pain: Involvement of neural reward systems. *PLoS One*, *5*(10), e13309. https://doi.org/10.1371/journal.pone.0013309

How to cite this article: Morato, C., Guerra, P., & Bublatzky, F. (2023). A partner's smile is not per se a safety signal: Psychophysiological response patterns to instructed threat and safety. *Psychophysiology*, *00*, e14273. https://doi.org/10.1111/psyp.14273