Shared safety abolishes the recovery of learned threat

Yafeng Pan[1], Andreas Olsson[1], & Armita Golkar[1,2*]

[1] *Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden*

2 Department of Psychology, Stockholm University, Stockholm, Sweden

[*] **Corresponding authors:**

Armita Golkar

armita.golkar@psychology.su.se

Abstract

Social learning offers an efficient route to transmit information about threat and safety. To better isolate the processes that contribute to the efficacy of social safety learning, we developed a novel dyadic model of associative threat and extinction learning. In three separate social groups, we manipulated whether safety information during extinction was acquired via direct exposure to the conditioned stimulus (CS) in the presence of another individual (Direct exposure), via observation of other's safety behavior (Vicarious exposure), or via the combination of both (Shared exposure).These groups were contrasted against a fourth group receiving direct CS exposure alone (Asocial exposure). Based on skin conductance responses, we observed that all social groups outperformed asocial learning in *inhibiting* the recovery of threat, but only Shared exposure *abolished* threat recovery. These results suggest that social safety learning is optimized by a combination of direct exposure and vicariously transmitted safety signals.

Keywords: social interaction, social learning, shared safety, recovery, threat

Statement of relevance

Humans, like other social animals, learn about threats and safety in the environment through social cues. Yet, the processes that contribute to the efficacy of social safety learning during threat transmission remain unknown. Here, we used a two-person approach to analyze skin conductance responses as participants engaged in a standard fear conditioning procedure (acquisition, extinction, and reinstatement). We found that during extinction, both (*i*) direct (conditioned stimulus) exposure in the presence of another individual and (*ii*) vicarious safety signals alone is sufficient to inhibit subsequent threat recovery, but that abolishing the recovery of conditioned threat responses requires a combination of both. This study has relevance for understanding how social information can optimize standard, asocial safety learning procedures to augment the effects of exposure on previously acquired fears. Thus, our work might help identify psychological and social strategies that can be used to counteract maladaptive fears in humans.

1. Introduction

Learning to predict and respond to threats and potential dangers in the environment is critical to adaption and survival. Such learning has been well described using associative threat learning models in both humans and animals (LeDoux, 2000), and has advanced our understanding of the etiology of threat and anxiety related disorder. These models are however inherently asocial as they rely on individual trial and error learning - neglecting that humans are inherently social. Most of us spend the majority of our daily lives socially connecting with others; these behaviors constitute social forms of learning, facilitating our acquisition of knowledge about the surrounding environment and influence our expectations and behaviors (Bandura, 1977; Olsson & Phelps, 2007).

Previous studies on social learning have mainly focused on how information about threats and potential dangers in the environment is transmitted between individuals (Debiec & Olsson, 2017; Lindström, Selbing, & Olsson, 2016) – a process known as *social threat learning* (LeDoux, 2017). Yet, less is known about how these learned threat responses can be changed and modified through socially acquired safety information. This is critical as deficits in the capacity to adapt and regulate responses to threatening cues when they no longer signal danger is a key feature of threat and anxiety-related disorders (Graham & Milad, 2011). Support for the utility of exploiting social information to regulate aversive responses mainly comes from research demonstrating that the presence of a conspecific attenuates the behavioral, physiological, and neural response to a stressful or threat-provoking event. This ”social buffering” phenomenon has been found across social species, including zebrafish (Faustino, Tacão-Monteiro, & Oliveira, 2017), rodents (Kiyokawa, Takeuchi, & Mori, 2007), non-human primates (Wittig et al., 2016), and humans (Ditzen & Heinrichs, 2014; Gunnar & Hostinar, 2015), suggesting the existence of an evolved system by social interaction can modulate behavior.

More specifically, previous experimental work in humans have demonstrated that acute physical (Roberts, Klatzkin, & Mechlin, 2015) and social stress (Heinrichs, Baumgartner, Kirschbaum, & Ehlert, 2003) result in significantly smaller elevations in heart rate, blood pressure, and cortisol in humans socially supported by another partner, compared to individuals who undergo stress alone. More recently, the mere presence of another partner was shown to reduce psychophysiological responses to aversive sounds, compared to an alone condition (Qi et al., 2020). The social buffering research has however almost exclusively focused on the social regulation of stress responses (Hostinar, Sullivan, & Gunnar, 2014) and little work has focused on how social information modulates learned threat memories or how social safety information is learned (but see also Ferreira et al., 2019). Briefly, extinction learning involves the formation of novel, safety associations after repeated safe exposures to a threatening stimulus; i.e. learning that the stimulus that earlier predicted danger now predicts safety. However, unlike threat learning, the threat-reducing effects of extinction learning decline with the lapse of time, do not generalize across different contexts, and are easily reinstated in behavior following re-exposure to cues associated with the original fear memory (Bouton, 2002). This recovery of threat responses after extinction also contributes to relapse of fear that is common after extinction-based exposure therapies that are used in the treatment of threat and anxiety-related disorders (Rachman, 1977). Consequently, one major aim of the research on threat learning and regulation is to use experimental extinction as a translational model to identify strategies that can prevent the recovery of threat after extinction in order to guide novel treatment approaches (Bouton, Mineka, & Barlow, 2001).

To address whether social information can optimize the effects of extinction learning and prevent the recovery of threat in humans, previous studies on both social threat and safety learning have commonly adopted video-based social learning paradigms (Golkar, Haaker, Selbing, & Olsson, 2016; Golkar, Selbing, Flygare, Ohman, & Olsson, 2013; Haaker, Golkar,

Selbing, & Olsson, 2017). In these paradigms, the observer learns about threat or safety by watching a pre-recorded video depicting the behavior of a demonstrator reacting either fearfully or calmly in the presence of threat. Although the recent decade has seen a paradigm shift towards real-time interactions in both social psychological and neuroscientific research (Pan et al., 2020; Redcay & Schilbach, 2019), few studies in humans have addresses whether social learning of threat and safety generalizes to a more ecologically valid setting. This methodological limitation critically also constrains the possibility to isolate core social aspects of the learning situation that can unveil the processes that contribute to safety transmission between individuals.

In the current study, we developed a dyadic experimental model with the aim to isolate the social processes that contribute to safety transmission. Given that extinction learning serves as a translational model for threat and anxiety disorders, we focused particularly on how social safety learning can prevent the recovery of learned threat responses. Live pairs of participants underwent an associative threat learning and extinction procedure followed by a threat recovery test. To isolate the social processes that are necessary for the efficacy of social safety learning, we ran three separate social safety learning groups: we manipulated whether safety information was acquired via own conditioned stimulus (CS) exposure in the mere presence of another individual (*Direct exposure*), via observation of another individual's safety behavior (*Vicarious exposure*) or via the combination of both (*Shared exposure*). These social safety learning groups were contrasted with a standard, asocial safety learning group who received direct CS exposure alone (*Asocial exposure*). Learning was indexed using skin conductance responses (SCRs) throughout all stages.

Our aim was to test two main predictions. First, to the extent that social buffering theory (Ditzen & Heinrichs, 2014; Gunnar & Hostinar, 2015) holds true, we would predict social safety learning (Direct, Vicarious, or Shared exposure) to outperform asocial safety learning

in *inhibiting* the recovery of threat. Second, in accordance with previous literature on safety learning (Golkar & Olsson, 2016; Golkar, Selbing, et al., 2013), we predicted that Shared exposure, involving both CS and model exposure, would *abolish* threat recovery. For this second hypothesis, we restricted our analysis to the early trials of reinstatement because previous research showing that reinstated threat responses typically undergo rapid re-extinction (Haaker, Golkar, Hermans, & Lonsdorf, 2014). Finally, based on previous research showing that inter-personal physiological synchrony is associated with social behaviors (Palumbo et al., 2017), including threat learning (Pärnamets, Espinosa, & Olsson, 2020), we carried out an exploratory analysis to assess the physiological synchrony between two partners in a dyad during social safety learning.

2. Methods

2.1. Ethics statement

This study was conducted in accordance with guidelines and regulations laid down in the Declaration of Helsinki. Written informed consent was obtained from each participant prior to the experiment. The study procedure was approved by Regional Ethical Review Board in Stockholm (DNR: 2016/249-31/1). Each participant was reimbursed with two movie tickets for their participation.

2.2. Participants

Power calculations were performed based on the obtained effect size (d = 0.75) in our original study comparing social and asocial extinction learning (Golkar, Selbing, et al., 2013). Setting the threshold to 95% sensitivity (1-beta = 0.95) and a significance level of alpha = 0.05 yielded a required sample size of 40 in each group. In an endeavor to increase statistical power, we decided to oversample to reach a minimum of 50 participants in each experimental

group. To achieve this, 231 participants were recruited from the student population at Karolinska Institutet. Participants were screened from having previously partaken in conditioning experiments and we confirmed that the participants in each dyad did not already know each other prior to participating in the experiment. Of the 231 participants who completed the experiment, we excluded 1 participant who failed to show any measurable skin conductance response (< 0.02 microsiemens, μS) to any CS during the entire experiment. This left 230 participants in the final sample. These participants were randomly assigned to either one of the three different social learning groups (Direct exposure, Vicarious exposure, and Shared exposure) or the asocial learning control group (for group characteristics and demographic information, see **Table1**).

Table 1. Group characteristics and demographic information.

Group	CS exposure	Social information	N	Gender	Age (SD)
Social safety learning					
Direct exposure	Yes	No	62	36F/26M	25.10 (4.11)
Vicarious exposure	No	Yes	60	33F/27M	25.46 (4.35)
Shared exposure	Yes	Yes	59	41F/18M	25.58 (3.80)
Asocial safety learning	Yes	/	49	22F/27M	24.94 (4.09)

Note. CS, conditioned stimulus; N, number; F, females; M, males; SD, standard deviation.

2.3. Experimental set-up

Social safety learning. Participants were enrolled in pairs of two participants that did not report previous contact or familiarity with each other. Each pair were seated in a waiting area during and were given time to read and sign informed consent forms before participation.

Then, each participant underwent a standardized shock calibration alone, with the aim to individually adjust the level of shock intensity to a level that was highly uncomfortable but not painful. Next, they were seated next to each other in the experimental room separated by a moveable wall that occluded any visual information of the other participant. In addition, the monitors were equipped with privacy screen filters to ensure that the visual presentations were not visible to the other individual in each pair. Participants were also instructed they were not allowed to talk to each other during any parts of the experiment. They were also informed that there would be brief interruptions during which the experimenter would make some adjustments and that should refrain from asking questions during these breaks.

Prior to the experiment (see **Figure 1** for an overview of the experimental stages), all participants were allowed to familiarize with the stimuli. Participants were exposed to two non-reinforced presentations of two different stimuli (one spider image and one snake image) that would subsequently serve as the conditioned stimuli (CSs). The formal experiment consisted of the following three stages: acquisition, extinction, and reinstatement test.

During the acquisition stage, all participants underwent a standard threat acquisition protocol involving two CSs. For each participant, one CS predicted the delivery of a shock (the unconditioned stimulus; US) on 2/3 of the CS presentations, whereas the other CS (CS-) never predicted the US. This acquisition stage ended with an instruction screen: “The experimenter will not make some minor adjustment, please remain seated”.

During the subsequent extinction stage, the experimenter removed the movable wall between the participants and adjusted mirrors that allowed the participants to receive visual feedback from the face of the other participant. The two mirrors were placed in an angle so that the reflection of the other participant’s face did not require them to move their gaze away from the visual presentation on their screen. We manipulated safety transmission in three separate groups by changing both CS exposure and instructions. In the *Direct exposure* group,

participants received the following written instructions: “During the next stage, ONLY you will be exposed to same pictures as you saw during the preceding stage, and the person sitting next to you will be exposed to two novel pictures. In the *Vicarious exposure* group, participants were informed that “During the next stage, you will be exposed to a presentation of two novel pictures, and ONLY the person sitting next to you will be exposed to the same pictures as you were exposed to during the preceding stage” and the *Shared exposure* group were informed “During the next stage, BOTH you and the person sitting next to you will be exposed to the same pictures as during the preceding stage”. For the Direct and Shared exposure groups, the extinction stage consisted of non-reinforced presentations of each CS. In the Vicarious exposure group, the CSs were replaced by the presentation of two neutral and non-related control stimuli (picture of a butterfly and a bird).

During the final reinstatement stage, the experimenter re-positioned the wall between the participants and removed the mirrors. After this readjustment, all participant received identical instruction on their screen “The experiment will now be resumed” which was followed by a standard *reinstatement* test during which participants–received 3 un-signaled shocks followed by a black screen (34 s) before all CSs were represented to the participants.

Asocial safety learning. Participants underwent the same experimental stages (acquisition, extinction, and reinstatement) using the same set-up but alone in the absence of another participant. After the acquisition stage, participants in this *Asocial exposure* group received the following instructions “During the next stage, you will now be exposed to the same two pictures as you were during the preceding stage”. This stage was followed by the reinstatement stage.

During all stages of the experiment, each CS was presented six times with a duration of 6 s. Between each CS trial, a black screen (i.e., inter-stimulus interval, ITI) was presented for 12–18 s.

Manipulation check. All pairs reported seeing the other participants face in the mirrors during extinction, without having to divert gaze from the CS presentation, and none of the included participants reported being able to see the CS presentation of the other individual.

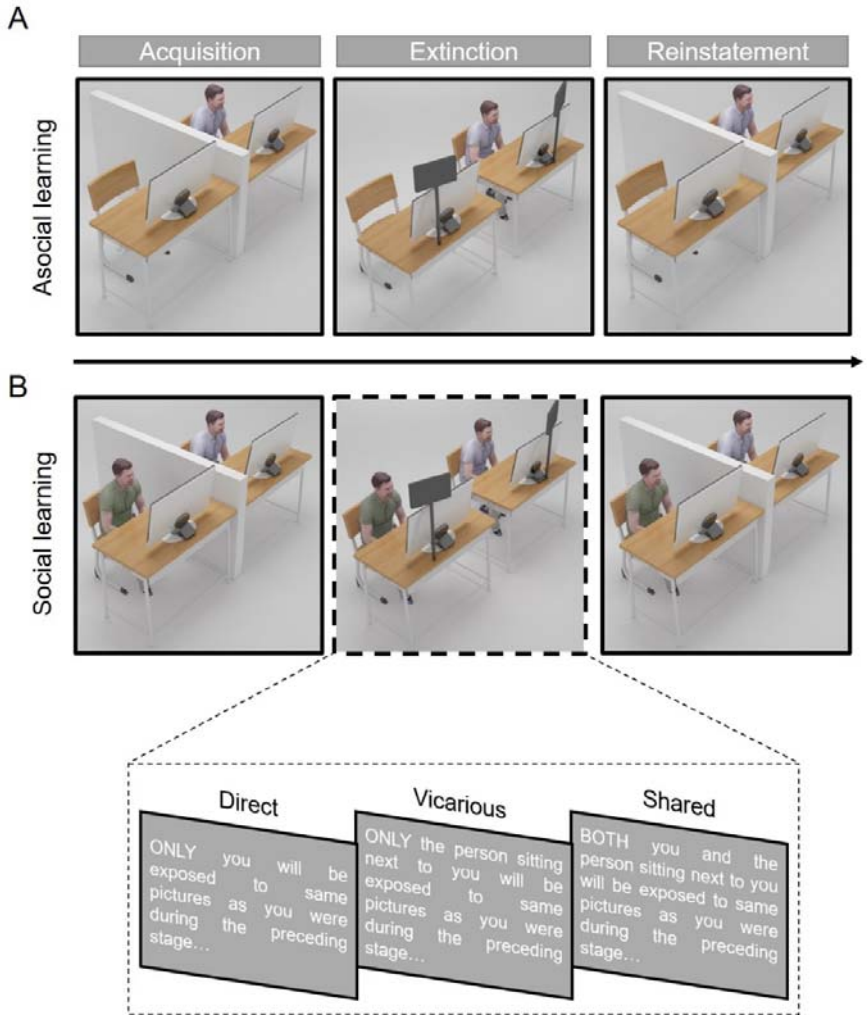

Figure 1. General set-up. All participants in (A) and (B) underwent an identical acquisition and reinstatement stage with the exception that in (A), the Asocial learning group performed all stages alone. (B) Participants in the social learning group were further categorized into three sub-groups (Direct, Vicarious, and Shared exposure) based on different CS exposure and instructions they received during extinction. Participants in both the Direct exposure and Shared exposure groups viewed the same CS as during the previous stage but either instructed that the person sitting next to them were observing novel, irrelevant stimuli (Direct exposure) or the same stimuli (Shared exposure) as themselves. Participants in the Vicarious exposure group were exposed to two novel,

neutral stimuli and instructed that only the person sitting next to them were exposed to the stimuli from the preceding stage. During all stages, screen filters occluding all information of what the other participant was exposed to were used.

2.4. Data acquisition

Participants' SCRs were recorded and measured using a Biopac MP150 system and AcqKnowledge 4.0 software (BIOPAC Systems Inc., Goleta, CA, USA). Ag-AgCl electrodes were attached to the middle and index fingers from the hand not receiving shocks. A conducting gel was used to improve signal quality, in accordance with standard recommendations (Boucsein, 2012). Electric shocks (aversive stimuli) were delivered using a STM200 module (BIOPAC Systems Inc., Goleta, CA, USA). The sampling rate was set to 2 kHz. Shocks consisted of a single 100ms DC pulse, which were applied to the lower forearm. The intensity of the shocks was individually calibrated, following the rule being "unpleasant but not painful". Stimulus presentation and shock delivery were controlled by the Presentation software (Neurobehavioral Systems, Inc., Albany, CA).

2.5. Data analysis

SCRs were continuously recorded throughout all experimental stages (acquisition, extinction, and reinstatement test). The raw signals were filtered offline with a low-pass filter (1 Hz) and a high-pass filter (0.05 Hz). The SCR was measured for each stimulus trial as the base-to-peak amplitude of the first response (in μS) in the latency window from 0.5 to 4.5 s following stimulus onset (Haaker et al., 2017). Minimal response criterion was set to 0.02 μS, and responses that did not pass this criterion were scored as 0. SCR scores were then *z*-transformed to normalize the distributions (Boucsein, 2012). Clean SCRs were then used for subsequent statistical comparisons.

In addition, to access the physiological synchrony between two participants in a dyad, clean SCR time series was further submitted to an exploratory analysis using the Pearson correlation coefficient (Piazza, Hasenfratz, Hasson, & Lew-Williams, 2020; Reddan, Young, Falkner, López-Solà, & Wager, 2020). The correlation coefficients between two SCRs from participant # 1 and participant #2 were calculated over the whole extinction stage (**Figure 1**). Resulting values were Fisher-z-transformed. Note that only Social exposure groups (Direct, Vicarious, and Shared exposure) were included in this exploratory analysis, as physiological synchrony should be estimated within dyads. One dyad containing a missing participant (failed in SCR acquisition, see section 2.2) were excluded, resulting in 31 pairs in the Direct exposure group, 30 pairs in the Vicarious exposure group, and 29 pairs in the Shared exposure group.

2.6. Statistical analyses

All the statistical analyses were conducted using IBM SPSS (version 21, SPSS Inc., Chicago, IL, USA). Each stage of the experiment was analyzed separately and significant interactions were followed up by separate two-tailed *t*-tests. Significance level was set to $p < 0.05$. η^2 and Cohen's *d* are reported as measures of effect size where appropriate.

2.6.1. Confirmatory analysis

When analyzing data from the acquisition stage, a 2 × 6 × 4 mixed-design Analysis of Variance (ANOVA), with Stimulus (CS+ and CS-) and Trial (#1~6) as two within-subject factors and Group (Direct, Vicarious, Shared, and Asocial exposure) as a between-subject factor, was used to analyze all results across conditions. Degrees of freedom were corrected using Greenhouse-Geisser estimates when assumption of sphericity was violated. We predicted successful threat acquisition in all experimental groups. Successful acquisition was defined as higher SCR to the CS+ relative to the CS- across acquisition, which would be

supported by a significant Stimulus × Trial interaction in the absence of any significant effects involving groups.

Data from the extinction stage were analyzed using the same mixed-design ANOVA. Extinction of the conditioned SCR was defined as decreasing SCR responses to the CS+ relative to the CS- across extinction training (Stimulus × Trial interaction). We anticipated a decreasing trend of CS+/CS- differentiation (as a marker of diminishing conditioned SCR) across trials in all groups except in the Vicarious exposure group that did not receive CS exposure during extinction and was therefore not expected to show significant CS+/CS- differentiation during this stage (no CS exposure, see **Table 1**).

2.6.2. Hypothesis-driven analysis

When analyzing data from the reinstatement stage, we carried out a series of statistics in a hypothesis-driven way. To test our first prediction i.e., that social safety learning would produce less recovery of learned threat compared to asocial safety learning, we compared CS+/CS- differentiation across reinstatement in a Stimulus (2) × Trial (6) × Group (4) mixed-design ANOVA, followed by planned comparisons between each social exposure group (Shared, Vicarious, or Direct) with the Asocial exposure group. To test our second hypothesis (i.e., that the Shared exposure group would abolish CS+/CS- differentiation during early reinstatement), we ran a Stimulus (2) × Trial (2: first 2 trials) × Group (4) mixed-design on SCR. We only focused on the early trials of reinstatement due to rapid re-extinction during reinstatement (Haaker et al., 2014), and in line with previous research demonstrating the efficacy of social safety learning in preventing the recovery of threat responses focusing on these early trials (Golkar, Castro, & Olsson, 2015; Golkar et al., 2016; Golkar, Tjaden, & Kindt, 2017). Each set of planned contrasts were Bonferroni corrected to account for multiple comparisons.

2.6.3. Exploratory analysis

Finally, we conducted exploratory analyses to test the potential physiological synchrony between two participants in a dyad under social settings (i.e. during extinction). Synchrony was estimated using the Pearson correlational approach, in accordance with recent advances (Piazza et al., 2020; Reddan et al., 2020). To evaluate the statistical significance of the Pearson correlation coefficients at the extinction stage, we created a null distribution based on a surrogate dataset. The surrogate dataset was obtained by Fourier transforming the original SCR signal, randomizing the phases of Fourier coefficients, performing an inverse Fourier transform, and re-computing the correlation coefficients (Liu et al., 2017). This procedure was repeated 1,000 times. Significance levels (thresholded at 0.05) were evaluated by comparing the correlation coefficients from the original dataset and those from the surrogate dataset. It was assumed that physiological synchrony based on original dataset should be higher than that based on surrogate dataset, as the latter disrupts the temporal structure of the original signals.

3. Results

3.1. Confirmatory results

3.1.1. Acquisition stage

First we confirmed that all group successfully acquired conditioned threat responses as supported by significantly higher SCR to the CS+ than the CS- across acquisition (Stimulus: $F_{1,226} = 13.74$, $p < 0.001$, $\eta_{partial}^2 = 0.06$; Stimulus × Trial interaction: $F_{4.71,1063.55} = 37.08$, $p < 0.001$, $\eta_{partial}^2 = 0.14$; see **Figure 2 (**left column**)**. There were no significant effects involving Group (main effect: $F_{3,\ 226} = 0.14$, $p = 0.94$; Trial × Group: $F_{12.93,974.11} = 1.17$, $p = 0.30$;

Stimulus × Group: $F_{3, 226} = 2.15$, $p = 0.10$; Stimulus × Trial × Group, $F_{14.12,1063.55} = 0.40$, $p = 0.98$).

3.1.2. Extinction stage

Second, we confirmed that CS+/CS- differentiation decreased across extinction. Progressive extinction was confirmed by a main effect of Trial ($F_{4.19,946.85} = 47.66$, $p < 0.001$, $\eta_{partial}^2 = 0.17$) and an interaction of Stimulus × Trial ($F_{4.58,1034.82} = 3.49$, $p = 0.01$, $\eta_{partial}^2 = 0.02$), with a generally decreasing trend in CS+/CS- differentiation (**Figure 2**, middle column). However, as expected, CS+/CS- differentiation differed between the Vicarious exposure group that did not receive own CS exposure and the other groups. This was evidence by a main effect of Group ($F_{3,226} = 3.87$, $p = 0.01$, $\eta_{partial}^2 = 0.05$) and a Stimulus × Group interaction ($F_{3,226} = 9.14$, $p < 0.001$, $\eta_{partial}^2 = 0.11$), indicating that CS+/CS- differentiation was significant in all groups (Asocial exposure, $t_{48} = 4.66$, $p < 0.001$, Cohen's $d = 2.14$; Direct exposure, $t_{61} = 4.61$. $p < 0.001$, Cohen's $d = 0.85$; and Shared exposure, $t_{58} = 6.39$, $p < 0.001$, Cohen's $d = 1.03$), except for the Vicarious exposure group that did not receive own CS exposure ($t_{59} = 0.42$, $p = 0.68$). This analysis did not yield a significant three-way interaction ($F_{13.74,1034.82} = 1.06$, $p = 0.39$).

To further elucidate the extinction effect, we analyzed CS+/CS- differentiation over the last two trial of the extinction stage. Results indicated that the Direct exposure ($F_{1,61} = 1.92$, $p = 0.17$) and Vicarious exposure ($F_{1,59} = 0.27$, $p = 0.61$) groups efficiently extinguished CS+/CS- differentiation, whereas the Shared exposure ($F_{1,58} = 6.12$, $p = 0.02$, $\eta_{partial}^2 = 0.10$) and Asocial exposure ($F_{1,48} = 5.40$, $p = 0.02$, $\eta_{partial}^2 = 0.10$) groups did not.

3.2. Hypothesis-driven results

To test our first prediction that all social safety learning groups would better inhibit reinstatement (i.e., less CS+/CS- differentiation) compared to the Asocial exposure group, we

ran a Stimulus (2) × Trial (6) × Group (4) mixed design ANOVA. Overall, this analysis yielded a significant Stimulus × Group interaction ($F_{3,226} = 3.40$, $p = 0.02$, $\eta_{partial}^2 = 0.04$), a main effect of Group ($F_{3,226} = 3.22$, $p = 0.02$, $\eta_{partial}^2 = 0.04$) and no interaction with Trial (Fs < 1.03, ps > 0.42). Comparisons of the separate social exposure groups (Shared, Vicarious, or Direct) vs. the Asocial exposure group using adjusted alpha levels of $p = 0.017$ (0.05/3) resulted in a significant Stimulus × Group interaction in the Shared vs. Asocial exposure comparison ($F_{1,106} = 7.64$, $p = 0.01$, $\eta_{partial}^2 = 0.07$), which, as predicted, was driven by lower CS+/CS- differentiation in Shared exposure relative to Asocial exposure (see **Figure 2**, right column). Corresponding analyses for the Vicarious vs. Asocial exposure group ($F_{1,107} = 5.08$, $p = 0.01$, $\eta_{partial}^2 = 0.05$) and the Direct vs. Asocial exposure ($F_{1,109} = 7.38$, $p = 0.01$, $\eta_{partial}^2 = 0.06$) were also significant at the adjusted alpha level.

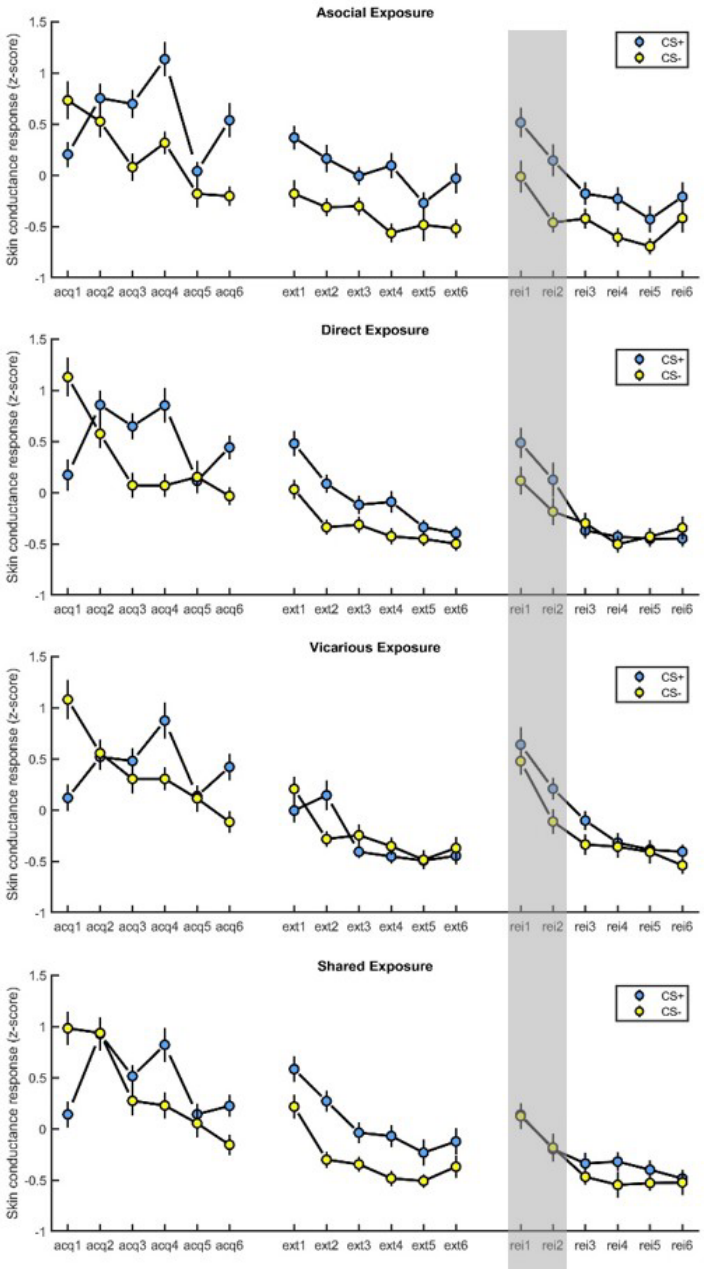

Figure 2. Mean skin conductance responses (z-score) across stimuli, groups, and trials over the entire experiment. The graphs from upper to bottom represent the data from Asocial exposure, Direct exposure, Vicarious exposure, and Shared exposure groups, respectively. Left column: acq1-acq6 refers to the 6 acquisition trials; middle column: ext1-ext6 refers to the 6 extinction trials; right column: rei1-rei6 refers to the 6 reinstatement trials. Grey shadow highlights the findings from the early reinstatement stage, indicating that shared safety abolishes the recovery of learned threat. Error bars denote standard errors.

To test our second prediction that the Shared exposure would abolish CS+/CS- differentiation during early reinstatement, we next analyzed early reinstatement test performance. This analysis resulted in a significant main effect of Group ($F_{3,226}$ = 3.64, p = 0.01, $\eta_{partial}^2$ = 0.05), and Trial ($F_{1,226}$ = 31.00, p < 0.001, $\eta_{partial}^2$ = 0.12) and a significant Stimulus × Group interaction ($F_{3,226}$ = 2.86, p = 0.036, $\eta_{partial}^2$ = 0.04). Using adjusted alpha levels of p = 0.017 (0.05/3), we observed the predicted difference between the Shared exposure and the Asocial exposure group (Stimulus × Group, $F_{1,106}$ = 9.19, p = 0.003, $\eta_{partial}^2$ = 0.08), and no significant differences between Vicarious vs. Asocial exposure ($F_{1,107}$ = 3.55, p = 0.06), or Direct vs. Asocial exposure ($F_{1,109}$ = 1.32, p = 0.25). Critically, in contrast to Asocial exposure (main effect of Stimulus, $F_{1,48}$= 23.21, p < 0.001, $\eta_{partial}^2$ = 0.33), Shared exposure abolished CS+/CS- differentiation during early reinstatement (main effect of Stimulus, $F_{1,58}$ = 0.00, p = 0.99, **Figure 3A**) using an adjusted alpha level of p = 0.025 (0.05/2). The absence of CS+/CS- differentiation during early reinstatement in the Shared Exposure group was further evidenced by a Bayesian paired samples *t*-test using JASP software (https://jasp-stats.org/, JASP Team, 2017), 95% Credible Interval = [-0.30, 0.21], Bayes Factor (BF_{10}) = 0.16 (**Figure 3B**).

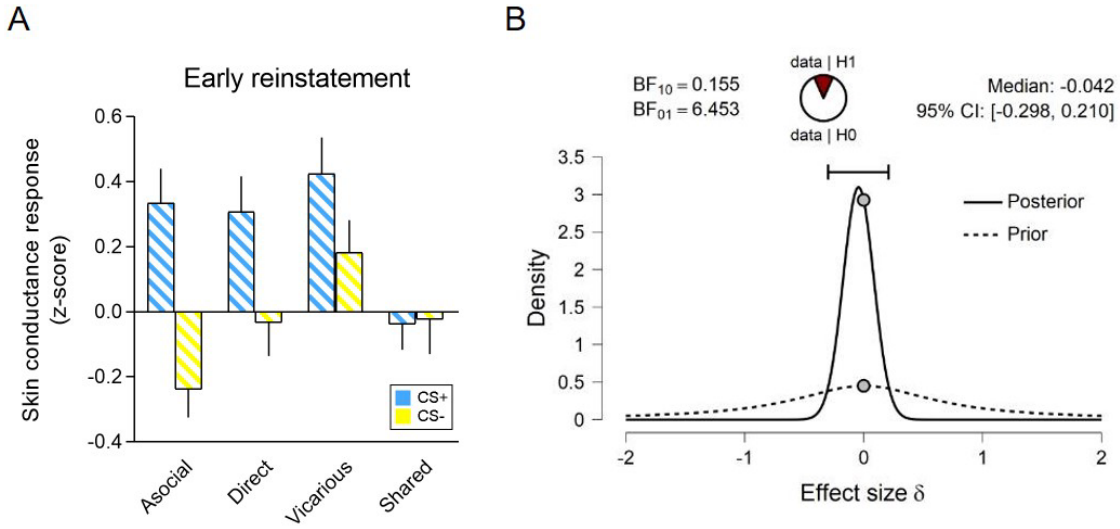

Figure 3. (A) Mean skin conductance response (z-score) across stimuli, groups during the early reinstatement stage. Error bars represent standard errors. (B) Inferential plots for prior and posterior probabilities. Skin conductance responses to CS+ and CS- during early reinstatement in the Shared Exposure group were compared using a Bayesian paired samples *t*-test.

3.3. Exploratory results

Our exploratory analysis focused on evaluating the physiological synchrony between two participants in a dyad at the extinction stage (**Figure 4**), during which the dyads were situated in a real-time social setting. All three social safety learning groups showed significantly stronger synchrony (averaged across trials and stimuli) between the SCR signals of two participants within dyads compared to a surrogate dataset (generated by phase-scrambling of original signals): Direct exposure (mean ± SD, 0.19 ± 0.05), Vicarious exposure (0.08 ± 0.05), and Shared exposure (0.20 ± 0.05), *p*s < 0.02.

An additional Stimulus (2) × Trial (6) × Group (3) mixed-design ANOVA on physiological synchrony revealed a main effect of Trial, $F_{4.46,388.12} = 7.66$, $p < 0.001$, $\eta_{partial}^2 = 0.08$, suggesting that synchrony generally attenuated across trials. The main effect of Stimulus was marginally significant, $F_{1,87} = 3.11$, $p = 0.08$, $\eta_{partial}^2 = 0.04$, indicating that synchrony was more pronounced for CS+ vs. CS-. No other significant effects were observed (*F*s < 1.91, *p*s > 0.15).

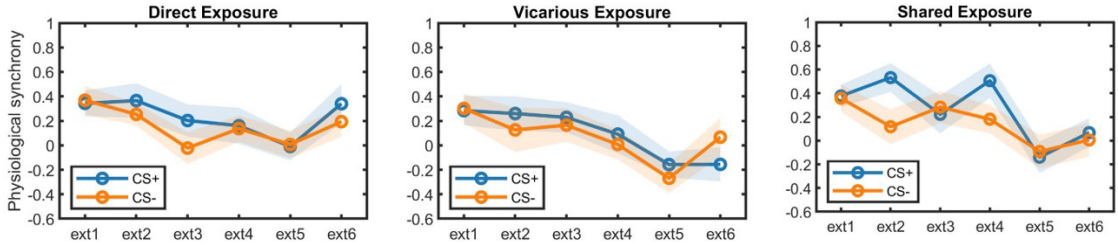

Figure 4. Physiological synchrony (average *r*) between skin conductance responses from two participants in a dyad across groups during the extinction stage. Ext1-ext6 refers to the 6 extinction trials. Shadow areas represent ± 1 standard errors.

4. Discussion

This study investigated how the recovery of learned threat responses are influenced by socially acquired safety information. In doing so, the study departed from two main hypotheses: (*i*) that social safety learning outperforms asocial safety learning in *inhibiting* the recovery of learned threat responses, and (*ii*) shared social safety learning *abolishes* threat recovery during early reinstatement. To address these hypotheses, we used a standard conditioning procedure (Golkar, Selbing, et al., 2013), which consists of three stages: (*i*) acquisition –threat acquisition through direct CS-US exposure; (*ii*) extinction –safety learning in the absence of aversive outcome (no US); and (*iii*) reinstatement – recovery of learned threat to the CSs after re-exposure to the US. Our results show that (*i*) compared to asocial safety learning, all three social safety learning groups were more successful in inhibiting the recovery of learned threat responses, as measured during a subsequent reinstatement stage, and (*ii*) only shared safety learning abolished the recovery of learned threat response, as evidenced by the non-significant CS+/CS- differentiation (confirmed by Bayesian factor analysis) during early reinstatement. These results suggest that both direct CS exposure in the presence of another person, and vicarious safety signals alone is sufficient to produce better inhibition of the recovery of learned threat responses compared to asocial learning, but that during early reinstatement, the beneficial effects of socially acquired safety relative to asocial safety learning are particularly pronounced when safety signals can be acquired from both direct experience and from another person (*Shared exposure*).

Previous studies supporting a modulatory role of social signals on aversive responses come mainly from non-human animal work (Cohen & Wills, 1985) and specifically from rodent models that have demonstrated that the expression of conditioned threat responses are reduced in the presence of a non-fearful conspecific (Ferreira et al., 2019). This previous

work indicates that a key factor facilitating safety learning is a social modulation mechanism, which requires the presence of a social conspecific and the trigger of safety signals during safety learning. Such “social modulation” might be further empirically characterized in terms of physiological synchrony (Palumbo et al., 2017). We observed higher synchrony for all social learning groups, compared to phase-scrambled controls. Thus, physiological synchrony might support shared experience and/or interpersonal information transfer by aligning physiological processes between the participants (Pärnamets et al., 2020). This study did not find any significant difference in physiological synchrony between the social groups; the lack of relationship deserves further investigations, as it may reflect either a lack of sensitivity regarding paradigms or analogous effects of social modulation on interpersonal autonomic coupling during safety learning.

The interpretation of our findings requires some considerations. First, the augmented inhibition of threat recovery during early reinstatement in the Shared exposure group does not seem to be related to end extinction performance. Neither Shared exposure nor Asocial exposure entirely abolished CS+/CS- differentiation during end extinction, but strikingly, they differed during the subsequent reinstatement stage. This lack of consistency between end extinction performance and recovery phenomena has been documented in previous research (Golkar, Bellander, & öhman, 2013; Prenoveau, Craske, Liao, & Ornitz, 2013). Second, an inspection of **Figure 3A** shows that whereas both the Shared and Vicarious extinction group showed relatively low CS+/CS- differentiation during early reinstatement, the lack of CS+/CS- differentiation in the Shared exposure group was due to comparably lower CS+ responses. In contrast, the Vicarious exposure group showed high CS+ *and* high CS- responses, indicating generalization of conditioned threat responses in this group. The stronger CS- response observed in the Vicarious exposure group might be related to the fact that this group did not receive CS exposure during the extinction stage and therefore showed

overall stronger SCRs to the re-appearance of the CSs. Third, the current study did not allow for verbal interaction between two individuals. However, to ensure that participants could access socially transmitted information, we set-up mirrors so participants could observe each other's emotional reactions (e.g., facial expression and body language) during the safety learning (i.e., extinction) stage. Fourth, the results from the exploratory analyses should be interpreted with caution, as our power calculations were based on a priori hypotheses for the main analyses of threat-related SCR responses. An inevitable consequence for the analyses focusing on physiological synchrony between dyads is that the sample size is reduced by 50% - which represents a common issue in dyadic studies (Pan & Cheng, 2020).

Taken together, these observations indicate that both direct CS exposure in the physical presence of another person, and the mere safety signal from another person alone is sufficient to inhibit the recovery of learned threat responses. Abolishing the recovery of threat, however, seems to require the combination of both (or beyond). As we did not have an a priori hypothesis for differences between the social safety learning groups (Direct vs. Vicarious vs. Shared exposure; i.e., which group best inhibits the recovery of learned threat), this question should be further addressed in future research powered enough to detect between-group differences. Nevertheless, to our knowledge, our work represents the first demonstration of social safety learning in a two-person situation with high ecological validity. The current set of findings extends recent research addressing the influence of physical presence on physiological responding to aversive stimuli (Qi et al., 2020) by addressing the influence of social modulation of threat in a safety learning context. More specifically, by focusing on safety learning, this work has relevance for understanding how social information can optimize standard, asocial safety learning procedure to augment the effects of exposure on previously acquired fears. Future research should more closely address the relative

contribution of different social processes to safety learning and delineate the underlying mechanistic features that characterize these processes.

Data accessibility

The full dataset used is accessible at https://osf.io/h3ckv/.

Authors' contributions

Armita Golkar conceived the study and designed the experiment with input from Andreas Olsson. Armita Golkar and Yafeng Pan performed the statistical analysis and interpreted the data and Yafeng Pan drafted the manuscript. All authors critically revised the manuscript and gave final approval for publication and agree to be held accountable for the work performed therein.

Acknowledgements

We would like to thank Lisa Espinosa, Cathelijn Tjaden and Anastasia Dima for help with the data collection and scoring. This research was supported by a Consolidator Grant (2018-00877) from the Swedish Research Foundation (Vetenskapsrådet), the Knut and Alice Wallenberg Foundation (KAW 2014.0237), and a Starting Grant (284366) from the European Research Council to Andreas Olsson, and an independent project grant to Armita Golkar from the Swedish Research Council (Vetenskapsrådet: 2017-02210).

Competing interests

The authors have declared that no competing interests exist.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change.

Psychological Review, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Boucsein, W. (2012). Principles of Electrodermal Phenomena. In *Electrodermal Activity* (pp. 1–86). https://doi.org/10.1007/978-1-4614-1126-0_1

Bouton, M. E. (2002, November 15). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, Vol. 52, pp. 976–986. https://doi.org/10.1016/S0006-3223(02)01546-9

Bouton, M. E., Mineka, S., & Barlow, D. H. (2001). A modern learning theory perspective on the etiology of panic disorder. *Psychological Review*, Vol. 108, pp. 4–32. https://doi.org/10.1037/0033-295X.108.1.4

Cohen, S., & Wills, T. A. (1985, September). Stress, Social Support, and the Buffering Hypothesis. *Psychological Bulletin*, Vol. 98, pp. 310–357. https://doi.org/10.1037/0033-2909.98.2.310

Debiec, J., & Olsson, A. (2017). Social Fear Learning: from Animal Models to Human Function. *Trends in Cognitive Sciences*, *21*(7), 546–555. https://doi.org/10.1016/j.tics.2017.04.010

Ditzen, B., & Heinrichs, M. (2014). Psychobiology of social support: The social dimension of stress buffering. *Restorative Neurology and Neuroscience*, *32*(1), 149–162. https://doi.org/10.3233/RNN-139008

Faustino, A. I., Tacão-Monteiro, A., & Oliveira, R. F. (2017). Mechanisms of social buffering of fear in zebrafish. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/srep44329

Ferreira, F. F., Rodrigues, F. S., Schmidt, S. D., Cavalcante, L. E. S., Zinn, C. G., Farias, C. P., … Izquierdo, I. (2019). Social support favors extinction and impairs acquisition of both short- and long-term contextual fear conditioning memory. *Neuroscience Letters*, *712*, 134505. https://doi.org/10.1016/j.neulet.2019.134505

Golkar, A., Bellander, M., & öhman, A. (2013). Temporal properties of fear extinction - Does

time matter? *Behavioral Neuroscience*, *127*(1), 59–69. https://doi.org/10.1037/a0030892

Golkar, A., Castro, V., & Olsson, A. (2015). Social learning of fear and safety is determined by the demonstrator's racial group. *Biology Letters*, *11*(1), 20140817. https://doi.org/10.1098/rsbl.2014.0817

Golkar, A., Haaker, J., Selbing, I., & Olsson, A. (2016). Neural signals of vicarious extinction learning. *Social Cognitive and Affective Neuroscience*, *11*(10), 1541–1549. https://doi.org/10.1093/scan/nsw068

Golkar, A., & Olsson, A. (2016). Immunization against social fear learning. *Journal of Experimental Psychology: General*, *145*(6), 665–671. https://doi.org/10.1037/xge0000173

Golkar, A., Selbing, I., Flygare, O., Ohman, A., & Olsson, A. (2013). Other people as means to a safe end: vicarious extinction blocks the return of learned fear. *Psychological Science*, *24*(11), 2182–2190. https://doi.org/10.1177/0956797613489890

Golkar, A., Tjaden, C., & Kindt, M. (2017). Vicarious extinction learning during reconsolidation neutralizes fear memory. *Behaviour Research and Therapy*, *92*, 87–93. https://doi.org/10.1016/j.brat.2017.02.004

Graham, B. M., & Milad, M. R. (2011). The Study of Fear Extinction: Implications for Anxiety Disorders. *American Journal of Psychiatry*, *168*(12), 1255–1265. https://doi.org/10.1176/appi.ajp.2011.11040557

Gunnar, M. R., & Hostinar, C. E. (2015). The social buffering of the hypothalamic–pituitary–adrenocortical axis in humans: Developmental and experiential determinants. *Social Neuroscience*, *10*(5), 479–488. https://doi.org/10.1080/17470919.2015.1070747

Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014, September 1). A review on human reinstatement studies: An overview and methodological challenges. *Learning and Memory*, Vol. 21, pp. 424–440. https://doi.org/10.1101/lm.036053.114

Haaker, J., Golkar, A., Selbing, I., & Olsson, A. (2017). Assessment of social transmission of threats in humans using observational fear conditioning. *Nature Protocols*, *12*(7), 1378–1386. https://doi.org/10.1038/nprot.2017.027

Heinrichs, M., Baumgartner, T., Kirschbaum, C., & Ehlert, U. (2003). Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biological Psychiatry*, *54*(12), 1389–1398. https://doi.org/10.1016/S0006-3223(03)00465-7

Hostinar, C. E., Sullivan, R. M., & Gunnar, M. R. (2014). Psychobiological mechanisms underlying the social buffering of the hypothalamic-pituitary-adrenocortical axis: A review of animal models and human studies across development. *Psychological Bulletin*, *140*(1), 256–282. https://doi.org/10.1037/a0032671

Kiyokawa, Y., Takeuchi, Y., & Mori, Y. (2007). Two types of social buffering differentially mitigate conditioned fear responses. *European Journal of Neuroscience*, *26*(12), 3606–3613. https://doi.org/10.1111/j.1460-9568.2007.05969.x

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience*, *23*(1), 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

LeDoux, J. E. (2017, May 1). Semantics, Surplus Meaning, and the Science of Fear. *Trends in Cognitive Sciences*, Vol. 21, pp. 303–306. https://doi.org/10.1016/j.tics.2017.02.004

Lindström, B., Selbing, I., & Olsson, A. (2016). Co-evolution of social learning and evolutionary preparedness in dangerous environments. *PLoS ONE*, *11*(8), e0160245. https://doi.org/10.1371/journal.pone.0160245

Liu, Y., Piazza, E. A., Simony, E., Shewokis, P. A., Onaral, B., Hasson, U., & Ayaz, H. (2017). Measuring speaker-listener neural coupling with functional near infrared spectroscopy. *Scientific Reports*, *7*, 43293. https://doi.org/10.1038/srep43293

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

1095–1102. https://doi.org/10.1038/nn1968

Palumbo, R. V, Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal Autonomic Physiology: A Systematic Review of the Literature. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *21*(2), 99–141. https://doi.org/10.1177/1088868316628405

Pan, Y., & Cheng, X. (2020). Two-Person Approaches to Studying Social Interaction in Psychiatry: Uses and Clinical Relevance. *Frontiers in Psychiatry*, *11*, 301. https://doi.org/10.3389/fpsyt.2020.00301

Pan, Y., Dikker, S., Goldstein, P., Zhu, Y., Yang, C., & Hu, Y. (2020). Instructor-learner brain coupling discriminates between instructional approaches and predicts learning. *NeuroImage*, *211*, 116657. https://doi.org/10.1016/j.neuroimage.2020.116657

Pärnamets, P., Espinosa, L., & Olsson, A. (2020). Physiological synchrony predicts observational threat learning in humans. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1927), 20192779. https://doi.org/10.1098/rspb.2019.2779

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2020). Infant and Adult Brains Are Coupled to the Dynamics of Natural Communication. *Psychological Science*, *31*(1), 6–17. https://doi.org/10.1177/0956797619878698

Prenoveau, J. M., Craske, M. G., Liao, B., & Ornitz, E. M. (2013). Human fear conditioning and extinction: Timing is everything...or is it? *Biological Psychology*, *92*(1), 59–68. https://doi.org/10.1016/j.biopsycho.2012.02.005

Qi, Y., Herrmann, M. J., Bell, L., Fackler, A., Han, S., Deckert, J., & Hein, G. (2020). The mere physical presence of another person reduces human autonomic responses to aversive sounds. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1919), 20192241. https://doi.org/10.1098/rspb.2019.2241

Rachman, S. (1977). The conditioning theory of fearacquisition: A critical examination. *Behaviour Research and Therapy*, *15*(5), 375–387. https://doi.org/10.1016/0005-7967(77)90041-9

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Reddan, M. C., Young, H., Falkner, J., López-Solà, M., & Wager, T. D. (2020). Touch and social support influence interpersonal synchrony and pain. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsaa048

Roberts, M. H., Klatzkin, R. R., & Mechlin, B. (2015). Social Support Attenuates Physiological Stress Responses and Experimental Pain Sensitivity to Cold Pressor Pain. *Annals of Behavioral Medicine*, *49*(4), 557–569. https://doi.org/10.1007/s12160-015-9686-3

Wittig, R. M., Crockford, C., Weltring, A., Langergraber, K. E., Deschner, T., & Zuberbühler, K. (2016). Social support reduces stress hormone levels in wild chimpanzees across stressful events and everyday affiliations. *Nature Communications*, *7*, 13361. https://doi.org/10.1038/ncomms13361