Fear in action: Fear conditioning and alleviation through body movements

Maria Alemany-González[1]*+, Martijn E. Wokke[1]*, Toshinori Chiba[2,3,4], Takuji Narumi[5], Katsumi Watanabe[6,7], Kimitaka Nakazawa[8], Hiroshi Imamizu[9,10,11], Ai Koizumi[1]+

1. Sony Computer Science Laboratories, Inc. Tokyo, Japan.
2. The Department of Decoded Neurofeedback, Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan
3. The Department of Psychiatry, Self-Defense Forces Hanshin Hospital, Kawanishi, Japan
4. The Department of Psychiatry, Kobe University Graduate School of Medicine, Kobe, Japan.
5. Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan.
6. Faculty of Science and Engineering, Waseda University, Tokyo, Japan.
7. Faculty of Arts, Design, and Architecture, University of New South Wales, Sydney, Australia.
8. Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
9. Research into Artifacts, Center for Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan.
10. Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, Tokyo, Japan.
11. Department of Cognitive Neuroscience, Cognitive Mechanisms Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan.
    (* equally contributed first authors, + corresponding authors)

Abstract

Learning a predictive cue of a threat - threat conditioning - allows us to prevent direct contact with the threat. In animals, threat-conditioned cues induce fear-like body movements such as freezing. Interestingly, training animals with alternative bodily defenses to avoid harm could reduce conditioned fear-like responses, suggesting that the physical capacity to actively defend against a threat could help overcome passively induced fear. However, evidence for the role of body movements in acquisition and alleviation of human fear is scarce. In the present study, human participants showed a deviation of multidimensional body movement patterns as threat conditioning progressed in a naturalistic virtual setting. The subsequent training to exert defensive body movements to prevent harm weakened the conditioned body movements and resulted in a long-term (24 hrs) reduction of the physiological responses and subjective fear toward the conditioned cues. Unlike the passive extinction procedure, the training of defensive body movements was resistant to spontaneous recovery of a fear response. The alleviating effect of the active defense training was observed when participants physically defended themselves but not when they merely observed another person vicariously defending against a threat on their behalf. These results suggest the critical role of body movements in both acquisition and reduction of conditioned fear among humans. Further investigations may leverage more embodied procedures to advance mechanistic understandings and clinical interventions of human fear memories.

Introduction

Experience of threats makes us learn the environmental cues that are predictive of the threats. This learning capacity - threat conditioning - enables animals and humans to minimize harm from threats as it can prevent direct contact with the threat upon perceiving its predictive cues (LeDoux, 2019; LeDoux, Moscarello, Sears, & Campese, 2017; Watson & Rayner, 1920). In the real world, minimization of harm requires bodily movements to act upon external situations such as attacking or escaping from the threats. Besides such well-documented fight-and-flight responses (Fanselow, 1988; Roelofs, 2017), animals show rich repertoires of defensive body movements to adaptively enhance survival in various threatening situations (Evans, Stempel, Vale, & Branco, 2019; Kozlowska, Walker, McLean, & Carrive, 2015; Ledoux & Daw, 2018; Mobbs et al., 2009; Wang et al., 2021). Accordingly, threat conditioning in animals is typically evaluated in terms of the emergence of defensive body movements in 3D spaces such as freezing (Bouton & Bolles, 1980; Brown, Kalish, & Farber, 1951; Davis, 1990). These suggest that optimization of defensive movement to better cope with specific threats is one ultimate purpose of threat association learning.

Interestingly, training defensive movement *per se* appears to result in reducing threat-conditioned responses. For instance, rats would show reduced conditioned response when they are trained to execute active avoidance behaviors, such as shuttering to the other side of a chamber to terminate a conditioned cue and threatening electric shock (Amat et al., 2005; Cain & LeDoux, 2007). The animal literature suggests that instrumental training of goal-oriented active avoidance could overcome more passive forms of fear-like responses (LeDoux et al., 2017). Different types of defensive responses are speculated to be executed through distinct neural pathways that are mostly connected with the amygdala (Ledoux & Daw, 2018). Learning to alternate from passive to active forms of defensive responses appears to be implemented by a prefrontal top-down facilitation of neural pathways supporting the latter (Ledoux & Daw, 2018; LeDoux et al., 2017; Wanke & Schwabe, 2020).

Unlike the demonstrated roles of body movements in both acquisition and reduction of threat association memories in animals, evidence in humans is scarce. This is despite the fact that tendencies for actions are considered to be central to human emotions, including fear (Arnold, 1960; Beckers, Krypotos, Boddez, Effting, & Kindt, 2013; Frijda, 1986; Lang, 1995) and its pathologies (Beckers et al., 2013). Instead, threat memories in humans are typically assessed with univariate physiological responses such as skin conductance responses (Steckle, 2010) and heart rates (Beckers et al., 2013; Cohen & Randall, 2003). Although those physiological responses may emerge as preparatory modulators of defensive bodily movements (Posada-Quintero, Florian, Orjuela-Cañón, & Chon, 2018; Rösler & Gamer, 2019; Trigili et al., 2019), they imply little about the role of overt bodily movements in human fear memories.

Moreover, although the active-avoidance procedure in animals (Amat et al., 2005; Cain & LeDoux, 2007) has been translated to human participants (Boeke, Moscarello, LeDoux, Phelps, & Hartley, 2017; Collins, Mendelsohn, Cain, & Schiller, 2014; Eldar, Hauser, Dayan, & Dolan, 2016; Hartley, Coelho, Boeke, Ramirez, & Phelps, 2019; Hartley, Gorun, Reddan, Ramirez, & Phelps, 2014;

Wanke & Schwabe, 2020), avoidance behaviors are typically implemented in a non-embodied manner. For example, participants make finger key-pressing responses to control icons on a monitor that are unrelated to the participants' bodily self. While some of the studies demonstrated the fear alleviating effect of active avoidance (Boeke et al., 2017; Delgado, Jou, LeDoux, & Phelps, 2009; Hartley et al., 2019, 2014; Wanke & Schwabe, 2020), the role of actual body movements awaits direct examination. Body movements are still well-required to defend against various real-life threats even in the modern human world - such as violent persons, looming vehicles, and various natural disasters. Thus, examining the efficacy of body movements in fear reduction should extend the translatability of the findings to real-life scenarios of fear.

The absence of bodily movements in human studies (Adolphs, 2013; De Gelder, 2006) partly owes to a common technical limitation. That is, experimental control needs to be achieved by constraining participants in a seated or laid posture, typically limiting the available actions to minimum peripheral movements such as in fingers and gaze. To overcome this limitation, we liberated body movements of human participants while taking advantage of the ecological validity of recent virtual reality (VR) techniques (Adolphs, 2013) to achieve threat conditioning in humans (Baas, Nugent, Lissek, Pine, & Grillon, 2004; L. Faul et al., 2020; Grillon, Baas, Cornwell, & Johnson, 2006; Reichenberger et al., 2019). With the measurements of naturalistic body movements, we aimed to directly investigate the roles of body movements in acquisition and reduction of fear in humans.

With VR, we presented one representative real-life traumatic scenario of interpersonal violence (Tada et al., 2015) where a specific person (CS+) acts out the threatening violence (US; unconditioned stimulus). We chose a scenario of interpersonal violence as it is one of the most common traumatic events in modern societies e.g., affecting 21-80% of adolescents (Finkelhor, Ormrod, & Turner, 2007; Kilpatrick et al., 2003), increases the risk for subsequent post-traumatic stress disorders (PTSD) (Cisler et al., 2012).

Experiment 1 was conducted across three consecutive days. On day 1, participants (N = 41) underwent threat conditioning in an unconstrained standing posture, where they were virtually intimidated and hit by each of the two male avatars (CS+s) but not by a third avatar (CS-) inside a virtual elevator (see Fig 1, Methods). On day 2, participants underwent a training session where they could actively defend themselves with a security spray to prevent violence from one of the CS+ avatars ($CS+_{active}$). The effect of the active-defense training was compared to an extinction procedure in the same VR setting. In the extinction training, participants were repetitively exposed to the other CS+ ($CS+_{extinction}$) avatar without further reinforcement by violence. On day 3, participants encountered the CS+ avatars without violence and we tested the spontaneous return of their threat-conditioned responses in the test (Fig 1A). Experiment 2 was conducted in an independent sample (N = 54) with similar procedures: We further examined whether fear could be alleviated only when the participants used their own body for active defense or even when they merely observed vicarious defensive movements by a third person. In both experiments, body movements of the participants were tracked and analyzed with representational similarity analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008; Lu & Ku, 2020; Visser, Kunze, Westhoff, Scholte, & Kindt, 2015), along with their physiological responses and subjective ratings.

Our active-defense training is a variant of, yet can be distinguished from, the previously used active-avoidance training in animal and human studies, by allowing participants to counter the occurrence of threat (US or violence). This reflects a unique feature of our training to *approach* and fight the CS+ to actively interrupt its capacity to inflict a threat (US or violence), which can be distinguished from the previous training that involves *avoidance* to flight or stay away from the CS+ and the subsequent US in rodents (Amat et al., 2005; Cain & LeDoux, 2007; LeDoux et al., 2017) and in humans (Boeke et al., 2017; Collins et al., 2014; Eldar et al., 2016; Hartley et al., 2019, 2014; Wanke & Schwabe, 2020). In other words, active-defense training assigns a participant with a role to break the external contingency between the CS+ and US, which assigns even more control over threatening situations to the participants than previous active avoidance procedures.

There were two main predictions. First, we predicted that threat conditioning results in the differentiation of multivariate body movement patterns between CS+ and CS- in human participants. Second, in line with previous studies demonstrating that active avoidance reduced conditioned responses more successfully compared to a passive extinction procedure (Boeke et al., 2017; Cain & LeDoux, 2007; Wanke & Schwabe, 2020), we predicted that active-defense (ACT) training would reduce threat-conditioned responses more effectively than the control extinction (EXT) training.

Anticipating the results, we found that the multidimensional body movements of human participants do alter as threat-conditioning progresses. We further found that training active-defense with inherent body movements can more effectively alleviate physiological fear response on a subsequent day, relative to the extinction training. The alleviating effect of the active defense training was observed when participants physically defended themselves but not when they merely observed another person vicariously defending against a threat on their behalf.

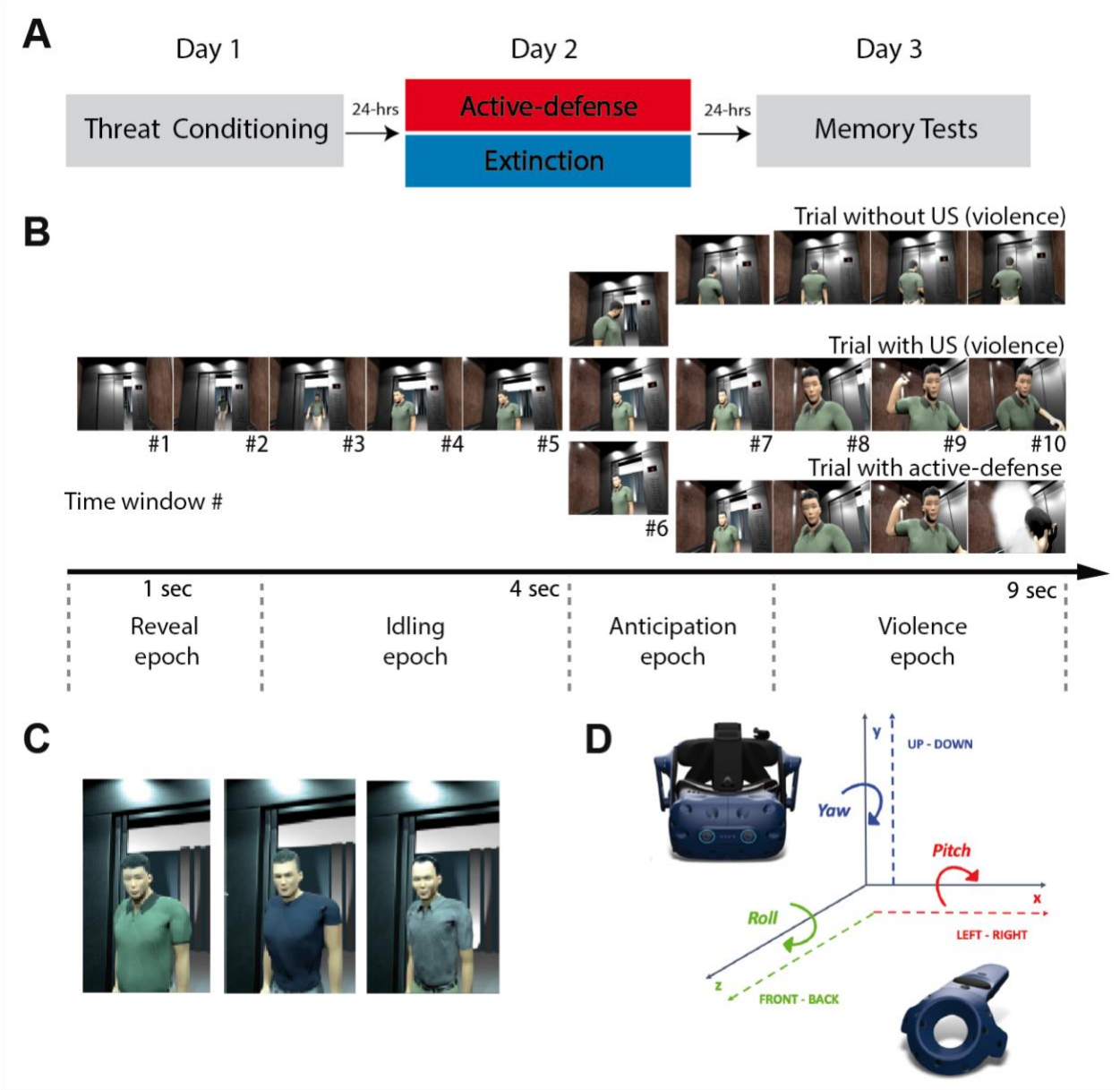

Figure 1. Schematics of the experimental design. (A) The flow of experimental sessions across three consecutive days. (B) Virtual scenes for non-violent trials (upper), violent trials (middle), and active-defense trials (lower) were presented across 10 representative time windows within a trial. (C) Three male avatars, two of which served as CS+s and another as CS- in a counterbalanced manner across participants. (D) Schematics of the three-dimensional spaces capture the participant's body movements as tracked in the head-mount-display and a hand-held controller.

Results

Threat conditioning in subjective and physiological measures

With a naturalistic VR threat-conditioning, participants successfully acquired conditioned responses (CR) when examined with their subjective ratings (Norrholm et al., 2006). After the conditioning session on day 1, participants reported higher feelings of fear as well as higher anticipation of unconditioned stimulus (US, violence) toward both $CS+_{active}$ and $CS+_{extinction}$ relative to CS- [one-tailed CS+ > CS-; Fear: $CS+_{active}$ $t(40) = 6.44$ $p < 0.001$; $BF_{+0} > 100$; $CS+_{extinction}$ $t(40) = 8.00$ $p < 0.001$; $BF_{+0} > 100$; US likelihood: $CS+_{active}$ $t(40) = 9.97$ $p < 0.001$; $BF_{+0} > 100$; $CS+_{extinction}$ $t(40) = 11.72$ $p < 0.001$; $BF_{+0} > 100$] (Fig 2A).

The CR with conventional skin conductance (SCR) emerged 24 hr following the conditioning session (Fig 2B), although it did not emerge during the conditioning session on day 1 (Fig S1A). On day 2, participants showed significant CRs (ΔSCR relative to CS-) in their initial encounter (first two trials) to both $CS+_{active}$ and $CS+_{extinction}$ avatars during the early phase of the training sessions [one-tailed CS+ > CS- ; $CS+_{active}$ $t(40) = 2.23$ $p = 0.01$; $BF_{+0} = 3.05$; $CS+_{extinction}$ $t(40) = 2.05$ $p = 0.02$; $BF_{+0} = 2.18$]. The CRs to $CS+_{active}$ and $CS+_{extinction}$ were statistically similar [$CS+_{active}$ Vs $CS+_{extinction}$ $t(40) = 0.35$ $p = 0.72$; $BF_{01} = 5.58$] (Fig 2B). Results remained quantitatively similar after removing outliers (3 SD apart from mean) (Supplementary Results and Fig S4A).

Threat conditioning in human body movements

As the conventional measures of subjective ratings and SCR suggested the success of threat conditioning, we went on to examine our first interest - whether participants show differential spatio-temporal patterns of body movement as they acquired threat association with the violence-predictive CS+ avatars. Specifically, we examined whether threat conditioning decreased the similarity of participants' body movement patterns upon encountering the CS+ avatars versus the CS- avatar with representational similarity analysis (RSA) (Kriegeskorte et al., 2008; Lu & Ku, 2020; Visser et al., 2015).

Body movement patterns were quantified with the position and rotation for the 3D axes of both head and right hand (12 dimensions), which were tracked with the VR head-mount-display and hand-held controller (Fig 1D). To obtain a measure that is independent from coordinate values and unbiased by previous positions during a trial, we calculated the absolute difference between two consecutive recorded samples (|Position(t2)-Position(t1)|) within trials, reflecting the magnitudes of movement per sample frame. To reduce the amount of comparisons while retaining enough temporal resolution to examine unique epochs of the scene (e.g. anticipation of violence), the spatial pattern similarity was evaluated for 10 time windows within each trial (9 sec) that captured *reveal (#1,2), idling (#3-5), anticipation (#6,7), and violence epochs (#8-10)* (see Fig 1B and methods), thus allowing us to investigate the temporal dynamics of spatial movement patterns.

We found that the similarity between CS+ and CS- decreased from the early to the late phase of the conditioning session (2 trials each). Strong evidence, as indicated by Bayes Factors (BF), was observed in the first and in the last three time windows within the trial, which correspond to the

periods of initial encountering with the CS avatars (*reveal-epoch*, see Fig 1B) and the period of potential violence occurrence (*violence-epoch*), respectively (time windows #1, 8, 9, & 10: all one-tailed *t*-values(40)> 3.26, *p*-values<0.001 FDR-corrected <0.05, Cohen's d > 0.51; all BF_{+0}>29.46, See Figure 2C for the results in other time windows). For time window #4 we observed a moderate effect [BF_{+0}=5.85], however when using frequentist statistics this effect did not survive the multiple comparison correction [*t*(40) = 2.56, *p* =0.007 FDR corrected>0.05, Cohen's d >0.39]. These results indicate that the movement patterns in response to CS+ avatars deviated more from the patterns in response to a CS- avatar as threat conditioning progressed with the CS+s.

The results were qualitatively similar when we examined the two CS+ avatars separately (see Fig S2). The similarity between the $CS+_{active}$ and CS- strongly decreased from the early to the late phase of the conditioning session in time windows #1 & 10 [all one-tailed *t*-values(40)> 2.97, *p*-values<0.002 FDR corrected<0.05, Cohen's d >0.464; BF_{+0}>14.92] and moderately decreased in time windows #4 & 8 [all one-tailed *t*-values(40)> 2.70, *p*-values<0.005 FDR corrected<0.05, Cohen's d >0.422; BF_{+0}>8.05]. All other times windows had BF_{+0}<1.11 (Fig S2A). We also observed similar results for the RSA between $CS+_{extinction}$ and CS- showing strong evidence in time windows #1, 9 & 10 [all one-tailed *t*-values(40)> 3.28, *p*-values<0.001 FDR corrected<0.05, Cohen's d >0.513; BF_{+0}>31.25] and moderate evidence in time windows #3, 5 & 8 [all one-tailed *t*-values(40)> 2.58, *p*-values<0.007 FDR corrected>0.05, Cohen's d >0.403; BF_{+0}>6.16]. All other times windows had BF_{+0}<1.20 (Fig S2A).

The results were specific to threat conditioning as they were not replicated when examining the deviation of body movement patterns between the two CS+s. In examining the changes in the similarity between $CS+_{active}$ and $CS+_{extinction}$ from the early to late phase of the conditioning session, we only found strong evidence of decreasing similarity for *the reveal-* and *idling-epoch* [time windows #1 & 5: two-tailed *t*-values(40)> 3.32, *p*-values<0.002 FDR corrected<0.05, Cohen's d > 0.51; BF_{10}>17.12]. Importantly, unlike the similarity of body movements between the CS+ versus CS-, the similarity between the two CS+s did not decrease in the later time windows when participants could start to anticipate potential violence in *anticipation* and *violence epochs* [time windows #6-10: two-tailed *t*-values(40)< 2.23, *p*-values>0.032 FDR corrected>0.05, Cohen's d < 0.35; BF_{10}<1.53]. These results suggest that a generic deviation of body movement patterns emerges at the start of a trial when an avatar's identity is revealed with the elevator door opening. Meanwhile, a deviation of body movement patterns due to threat conditioning may emerge specifically in the later *anticipation-* and *violent-epochs* of the trial.

To unpack the nature of threat-conditioned body movement patterns in response to the CS+s, we ran RSAs for head and hand separately. For head movement patterns we found differences between the phases of the conditioning session in time windows #1 & 10 [*t*-values(40)> 2.59, *p*-values<0.007 FDR corrected<0.05, Cohen's d > 0.40; BF_{+0}>6.23] (Fig S2B). For hand movement patterns we found differences between the phases of the conditioning session in time windows #4, 5, 8 & 9 [*t*-values(40)> 2.75, *p*-values<0.004 FDR corrected<0.05, Cohen's d > 0.43; BF_{+0}>8.92], suggesting that threat condition impacted more threat-relevant time windows of the hand than the head movement including *violence-epoch* (Fig S2B). When exploring RSAs ran with the position and rotation information for each axis - x (left-right), y (up-down), and z (front-back) - of the hand and head separately, we found no effect of phase of the conditioning session [all Fs<3.87, all ps>0.06] (Fig S2B). This may not be surprising as we lose access to the spatial

pattern information across axes by running RSA per axis. However, left-right movement of the hand was qualitatively distinguished between the trials encountering the CS+s avatars and the CS- avatar (see Fig S2B).

These results suggest that the spatio-temporal patterns of body movements alter as a consequence of threat conditioning in humans. The threat-conditioned movement patterns may partly depend on specific threatening scenarios and thus the currently observed patterns may not directly generalize to other scenarios. Nevertheless, the importance of those findings is that threat-conditioning results in specific qualitative changes of anticipatory body movement patterns opted for an experienced threat.

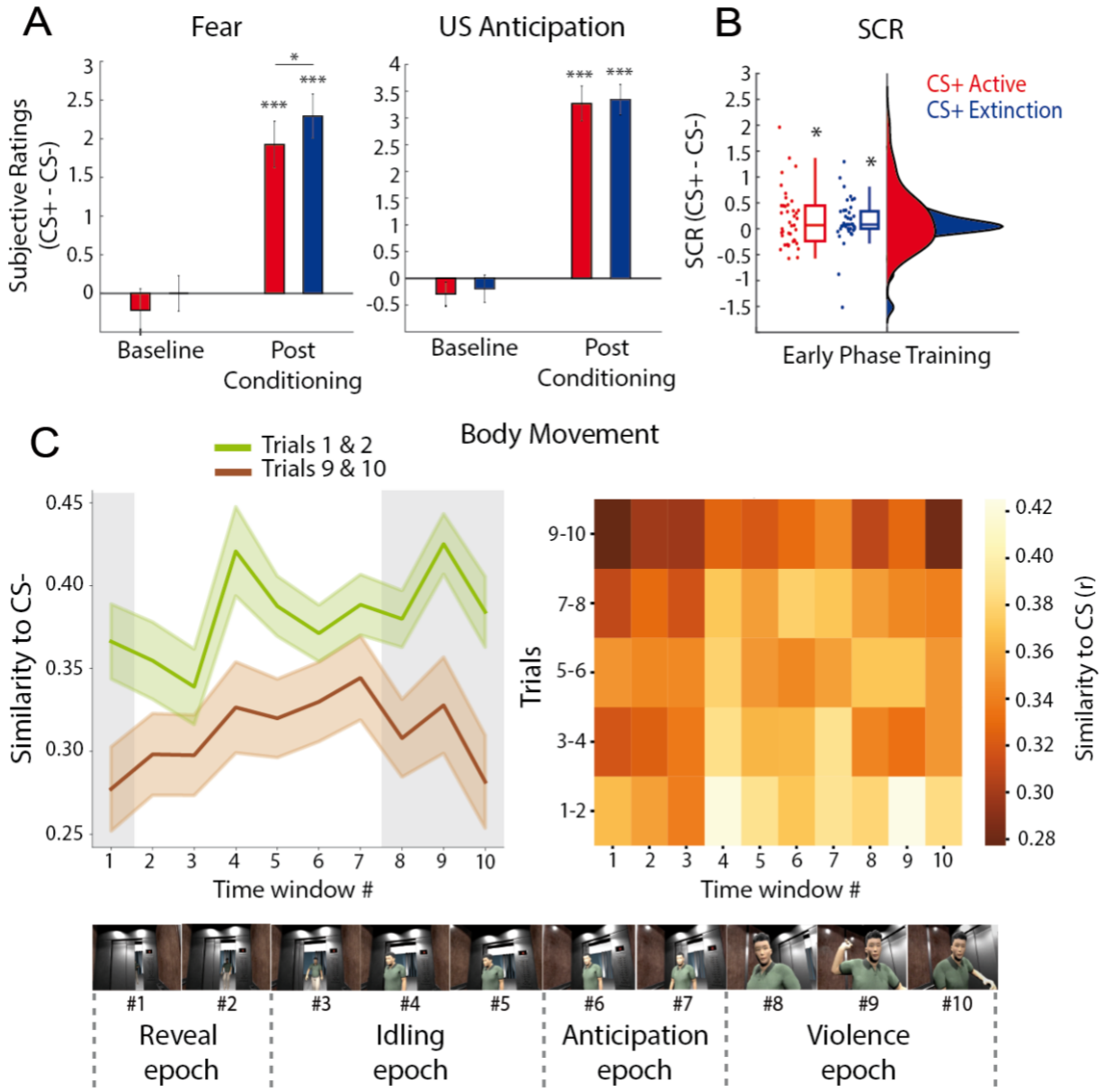

Figure 2. Threat conditioning in subjective ratings, SCR and body movements. (A) Conditioned responses [CS+ - CS-] in subjective ratings of fear and US-anticipation before and after conditioning on day 1. (B) Conditioned responses in SCR during the early phase (trials 1-2) of the training sessions on day 2. (C) Decreasing similarity in body movement patterns toward the CS+ versus the CS- avatars from early (trials 1-2) to late phase (trials 9-10) (left) or across all phases of the conditioning session (right).

Fear reduction effects through embodied *active-defense*

Next, we examined the immediate (day 2) and long-term (day 3) fear alleviating effects of active-defense (ACT) training conducted on day 2, which required the participants to physically defend against one of the two CS+ avatars ($CS+_{active}$) (Fig1B).

Immediate effects (day 2)

We first examined the immediate effects of the training on day 2. After the EXT training participants showed a significant reduction in subjective ratings of both fear and US-anticipation to the $CS+_{extinction}$ relative to the post-conditioning ratings on day 1 [day 1 Vs day 2; Fear: $t(40) = 5.25$, $p < 0.001$; $BF_{10} > 100$; US-likelihood: $t(40) = 9.48$, $p < 0.001$; $BF_{10} > 100$]. However, ACT training did not reduce subjective ratings relative to day 1 [day 1 Vs day 2; Fear: $t(40) = 0.88$, $p = 0.38$; $BF_{10} = 0.24$; US-likelihood: $t(40) = 1.33$, $p = 0.19$; $BF_{10} = 0.38$]. Subjective ratings to the $CS+_{active}$ were significantly higher than the $CS+_{extinction}$ on day 2 [$CS+_{active}$ Vs $CS+_{extinction}$ Fear: $t(40) = 3.63$, $p < 0.001$; $BF_{10} = 37.48$; US-likelihood: $t(40) = 7.42$, $p < 0.001$; $BF_{10} > 100$] (Fig 3A). These results indicate that only the EXT training reduced subjective fear and US-anticipation when tested immediately after the training on day 2. This is potentially due to further conditioning effects from the $CS+_{active}$ trials reinforced with the signs of threatening violence (i.e., the $CS+_{active}$ avatar raising his fist) in the ACT training session.

We next examined the immediate effects of the training on SCR on day 2. While there were significant CR to both CS+s in the early phase of the training session (see *Threat conditioning* in Results and Fig 2B), in the late phase of the training sessions, CR to both the $CS+_{extinction}$ and $CS+_{active}$ were no longer significant [$CS+_{extinction}$ $t(40) = -1.35$, $p = 0.18$; $BF_{10} = 0.39$; $CS+_{active}$ $t(40) = 1.52$, $p = 0.13$; $BF_{10} = 0.48$]. This suggests immediate fear reduction effects with both types of training. Nevertheless, CR to the $CS+_{active}$ was significantly higher than CR to the $CS+_{extinction}$ [$t(40) = 2.408$, $p = 0.02$; $BF_{10} = 4.31$], suggesting advantage of the EXT training in immediate fear reduction (Fig 3B). Similar to the aforementioned fear rating results, the relatively higher SCR with $CS+_{active}$ is likely due to the fact that the $CS+_{active}$ was further reinforced by the signs of threatening violence (i.e., the avatar raising his arm).

We did not examine the body movement patterns during the training sessions. This is because the task requirements to make specific movements of defense (i.e., spraying) made the movement patterns distinct during the active-defense training (Fig S3), rendering it difficult to meaningfully compare the patterns with the extinction training. Specifically, even from the start of the training (trials 1-2), similarity in body movement patterns toward the CS+ versus the CS- avatars deviated between the ACT and EXT condition in all time windows [all two-tailed *t*-values(40)> 2.96, *p*-values<0.005] (Fig S3A), supporting that different task requirements (i.e., using spray only in the ACT training) confounded the analyses of movement pattern during the training.

Long-term (24 hr) effects (day 3)

While the EXT training generally showed an advantage in its immediate fear reduction effect on day 2, the ACT training outperformed the EXT training in its long-term (24 hr) effect tested on day 3 especially with SCR.

On day 3, subjective ratings of fear to $CS+_{active}$ and $CS+_{extinction}$ were no longer different [$CS+_{active}$ Vs $CS+_{extinction}$ $t(40) = 1.27$, $p = 0.21$; $BF_{10} = 0.35$]. This was due to that subjective rating of fear to the $CS+_{active}$ significantly decreased from day 2 to day 3 [Fear: post-tests day 3 Vs post-training day 2; $t(40) = 3.302$, $p = 0.002$; $BF_{10} = 16.2$] while the fear rating to the $CS+_{extinction}$ on day 3

remained equivalent to day 2 [$t(40) = -.0.96$, $p = 0.34$; $BF_{10} = 0.26$]. Meanwhile, subjective ratings of US-likelihood remained higher with $CS+_{active}$ than $CS+_{extinction}$ [$CS+_{active}$ Vs $CS+_{extinction}$ $t(40) = 1.10$, $p = 0.04$; $BF_{10} = 1.21$]. This was despite the fact that the rating of US-likelihood for $CS+_{active}$ was decreased on day 3 relative to day 2 [$t(40) = 6.89$, $p < 0.001$; $BF_{10} > 100$], suggesting the emergence of some long-term effect. There was no such reduction with $CS+_{extinction}$ from day 2 to day 3 [$t(40) = 0.23$, $p = 0.81$; $BF_{10} = 0.17$] (Fig 3A). These results suggest that ACT and EXT training had a comparable long-term reduction effect on subjective rating of fear, while ACT training maintained a higher subjective anticipation of violence on day 3.

Critically, we observed a more robust long-term reduction of SCR to $CS+_{active}$ relative to $CS+_{extinction}$. In the early phase of the spontaneous recovery test, a significant CR to the $CS+_{extinction}$ reemerged after 24hrs through spontaneous recovery [$t(40) = 2.21$, $p = 0.03$; $BF_{10} = 1.49$], return of extinguished fear-like response with the passage of time (KM Myers & M Davis 2007 Mol Psychiatry). However, $CS+_{active}$ did not elicit a CR [$t(40) = 0.76$, $p = 0.45$; $BF_{01} = 4.51$] (Fig 3B). This return of response since the end of the training was quantified with the spontaneous recovery Index (SRI), or change of CR from the late phase of the training on day 2 (immediate effects) to the early phase of the test on day 3 (long-term effects) similarly to a previous study (EA Boeke et al, 2017). The SRI to $CS+_{extinction}$ was significantly higher than the SRI to $CS+_{active}$ [$t(40) = -2.57$, $p = 0.007$; $BF_{+0} = 6.13$] (Fig 3B). Together, the fear reduction effect of ACT training may not emerge immediately, as indicated by relatively higher response to $CS+_{active}$ by the end of the training. Nevertheless, over a day, the ACT training induced a more robust long-term fear reduction effect than the EXT training and prevented spontaneous recovery. Results remained similar after removing outliers (Supplementary Results and Fig S4A).

We further examined whether the EXT and ACT training had differential effects on body movement patterns on day 3. Specifically, we examined whether the threat-conditioned body movement patterns with CS+s became similar to the CS- again after the training. In the early phase of the spontaneous secovery test on day 3, we observed moderate evidence (as indicated by the Bayes factors) that body movement patterns with CS+ became more similar to CS- during *anticipation-epoch* after the ACT training than after the EXT training although the differences between $CS+_{active}$ and $CS+_{extinction}$ did not survive multiple comparisons correction with a frequentist statistical approach [time windows #6 & 7: $t(40) = 2.53$, $p = 0.008$ FDR corrected >0.05, Cohen's d =0.40, BF_{+0}=5.46; $t(40) = 2.26$, $p = 0.015$ FDR corrected >0.05, Cohen's d =0.35, BF_{+0}=3.20; for all other times windows BF_{+0}<1.01], see Figure 3C. The body movement patterns of $CS+_{active}$, became more similar to CS- compared to the similarity of $CS+_{extinction}$ to CS-, especially in *anticipation-epoch* across time windows #6 & 7 of head movement [$t(40) = 3.10$, $p = 0.002$ FDR corrected < 0.05, Cohen's d = 0.48, $BF_{+0} = 19.58$; $t(40) = 2.83$, $p = 0.004$ FDR corrected <0.05, Cohen's d =0.44, BF_{+0}=10.78; for all other times windows BF_{+0}<0.89]. For hand movement, we did not find a significant difference between the CS+ avatars [all *t*-values<2.02, all *p-values*>0.025 FDR corrected >0.05, all BF_{+0}<2.05]. The observed movement patterns on day 3 appear to deviate from the ones that particularly separated the movements for the CS+s versus CS- during the threat-conditioning session (i.e., movement of the hand rather than the head). Thus, the ACT training may lead to adopting novel defense movement patterns rather than merely reversing the body movement patterns to the original non-threat-conditioned state.

Effects on fear renewal and reconditioning (day 3)

We further tested the CR when participants encountered the CS avatars in a novel context in the renewal test session (see Methods). However, the comparison between the two CS+s was difficult as their CRs were already extinguished with the unreinforced trials during the first test which preceded the Renewal test (see Supplementary Results and Fig S1A). Potentially because of these additional extinction effects from the spontaneous recovery test, the results of the subsequent reconditioning session were only suggestive (see Methods, Supplementary Results, and Fig S1A).

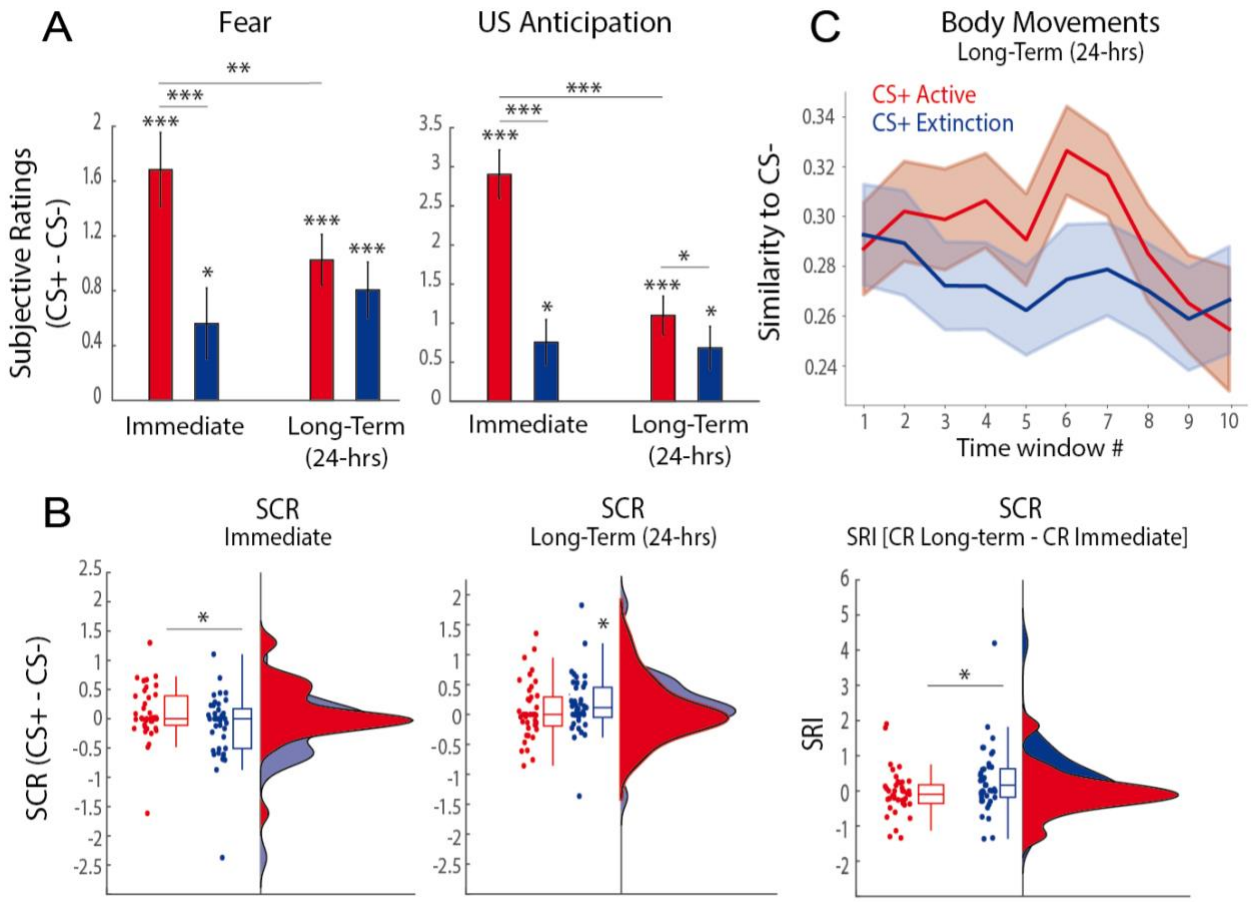

Figure 3. Comparison of fear alleviation effects between Active-defense (ACT) training and Extinction (EXT) training. (A) Immediate (day 2) and long-term (day 3) effects of ACT and EXT training on conditioned responses [CS+ - CS-] in subjective ratings of fear and US-anticipation. (B) Immediate and long-term effects of ACT and EXT training on conditioned SCR (left and middle) and Spontaneous Recovery Index (SRI), as measured by the difference between the immediate versus the long-term conditioned SCR (right). (C) Long-term effects of ACT and EXT training on body movements patterns. Bayesian factor indicates moderate evidence for a greater similarity in body movement patterns between $CS+_{active}$ and CS- relative to the similarity between $CS+_{extinction}$ and CS-.

Fear reduction effects requires self *active-defense*

Finally, we examined whether the fear reduction effect of ACT training requires one's own body movements. To this aim, a different group of 54 participants underwent Experiment 2 where they were assigned to one of the two ACT training groups on day 1: One group underwent a Self ACT training that consisted of the similar procedure as in Experiment 1, where they actively defended themselves with a virtual security spray to prevent violence. Another group underwent a Vicarious ACT training that consisted of mere observation of a third avatar defending with a security spray on behalf of the participants. The trial-wise timing of defensive movement (spraying) by the third avatar in a given participant from the Vicarious ACT group was matched to that of the paired participant from the Self ACT group. To avoid a possible confound of social buffering due to the accompaniment of a third person during fearful experience (Eisenberger et al., 2011) [42], the same third avatar was also present during the Self ACT training.

Besides the introduction of the two ACT training types, the overall procedure was similar to Experiment 1 with three exceptions. First, since the aim of Experiment 2 was to compare the fear responses between the groups, we did not include the EXT training as a reference. Instead, the participants underwent the ACT training with one of the two avatars ($CS+_{active}$) while the other CS+ received no training ($CS+_{untrained}$) (Fig 4A). Second, both Threat conditioning and ACT training sessions were conducted on day 1. Third, on day 2 participants underwent two consecutive spontaneous recovery tests: In the first test participants encountered the CS avatar alone as in Experiment 1 (Test-Alone) and in the second test the CS avatar was accompanied by the same third avatar from the Vicarious ACT training (Test-Accompanied) (Fig 4A). See supplementary for the full description of the method of Experiment 2.

Subjective Ratings

Both Self and Vicarious ACT groups reported higher feelings of fear as well as higher anticipation of violence toward both $CS+_{active}$ and $CS+_{untrained}$ relative to CS- after threat conditioning on day 1, indicating successful threat conditioning (Fig S5). Next, we examined the immediate and long-term fear alleviating effects of the two types of ACT training on day 1 and day 2, respectively.

Similarlty to ACT training effects in Experiment 1, both self and vicarious ACT training showed immediate and long-term reduction of subjective fear ratings while only long-term reduction of US-expectancy. While both ACT training effects on subjective ratings were generally equivalent to no-training effects, vicarious ACT training showed higher sujective fear and US-expectancy ratings relative to non-training (Fig S5). See supplementary for the full description of the rating results of Experiment 2.

SCR

We next examined the effects of Self and Vicarious ACT training on SCR. Note that, as $CS+_{untrained}$ was absent during the training session, we did not compare immediate effects (late training) and SRI (relative to late training) between $CS+_{active}$ and $CS+_{untrained}$.

SCR toward both CS+ avatars was statistically similar between the two groups before the ACT training session (see Supplementary results of Experiment 2 - SCR section). Although CR was

not significant during the conditioning session nor during the early ACT training session, this result is consistent with Experiment 1 which showed that CR emerges only from day 2. Thus, we went onto assess whether the Self and Vicarious ACT training differed in their long-term fear alleviation effect on day 2 relative to no-training, using repeated measures ANOVA. In the Test-Alone we found a significant interaction between ACT training type and CS+ type [$F(1,50) = 5.03$ $p = 0.02$] while in the Test-Accompanied we did not [$F(1,50) = 2.27$ $p = 0.13$]. In the Test-Alone, only the Self ACT training significantly reduced CR toward $CS+_{active}$ relative to the $CS+_{untrained}$ [$t(25) = -1.91$, $p = 0.03$; $BF_{10} = 1.91$] while the Vicarious ACT training did not [$t(25) = 1.25$, $p = 0.22$; $BF_{10} = 0.47$]. Only the Vicarious ACT training elicited a numerical CR toward $CS+_{active}$ (Vs CS-) [$t(25) = 1.96$, $p = 0.06$; $BF_{10} = 1.08$] while the Self ACT training did not [$t(25) = 1.06$, $p = 0.29$; $BF_{10} = 0.03$]. These results suggest that only the Self ACT training showed long-term fear alleviation effects in the Test-Alone relative to no-training. Interestingly, the Vicarious ACT training elicited a strong immediate fear response in the late training session, showing lower SCR towards $CS+_{active}$ relative to CS- [$t(25) = -2.23$, $p = 0.03$; $BF_{10} = 1.70$] but 24 hours later CR reemerged in the Test-Alone. SRI after Self ACT training was significantly lower than SRI elicited by Vicarious ACT training [unpaired t test; $t(52) = -2.20$, $p = 0.03$; $BF_{+0} = 1.97$] (Fig 4B). These results suggest that only the Self ACT training showed long-term fear reduction effects.

Body Movements

Finally we examined the conditioning and the training effect on body movement patterns. Replicating Experiment 1, we found strong evidence in both groups of decreasing similarity between CS+ and CS- including *anticipation-epoch* and/or *violence-epoch* from the early to the late phase of the conditioning session (2 trials each) [Self ACT: time windows #7 & 10: all one-tailed t-values(27)> 4.74, p-values<0.001 FDR-corrected <0.05, Cohen's d > 0.93; all BF_{+0} > 139.6 ; Vicarious ACT: time windows #3 & 8: all one-tailed t-values(27)> 3.81, p-values=0.005 FDR-corrected <0.05, Cohen's d > 0.56; all BF_{+0} > 4.65] (Fig 4D). These results indicate that the movement patterns captured that CS+ avatars deviated more from the CS- avatar as threat conditioning progressed.

We next examined whether the Self and Vicarious ACT training had differential effects on body movement patterns on day 2. Specifically, we examined whether the threat-conditioned body movement patterns with CS+s became similar to the CS- again after the training. We observed no differences in the CS- similarity patterns between $CS+_{active}$ and $CS+_{no\text{-}training}$ in any ACT training group in the Test-Alone. However, in the Test-Accompanied, we found different effects of the Self and Vicarious ACT training groups. In the Self ACT training group we observed anecdotal evidence of greater similarity to CS- with $CS+_{active}$ compared with $CS+_{no\text{-}trained}$ in the second and last time window of the trial in *reveal-* and *violence-epochs* [time window #2 & 10 $t(26) > 2.48$, $p < 0.02$ FDR corrected >0.05, Cohen's d > 0.48, $BF_{+0} > 2.65$], while in the Vicarious ACT training group we observed anecdotal evidence of greater similarity to CS- only in the last time window of *violence-epoch in* the trial [time window #10: $t(26) > 2.48$, $p = 0.02$ FDR corrected >0.05, Cohen's d = 0.48, $BF_{+0} = 2.44$] (see Fig 4C). When using frequentist statistics these effects did not survive the multiple comparison correction. Although suggestive with anecdotal evidence, these results suggest that Self ACT training might result in better long-term attenuation of threat-conditioned body movement patterns.

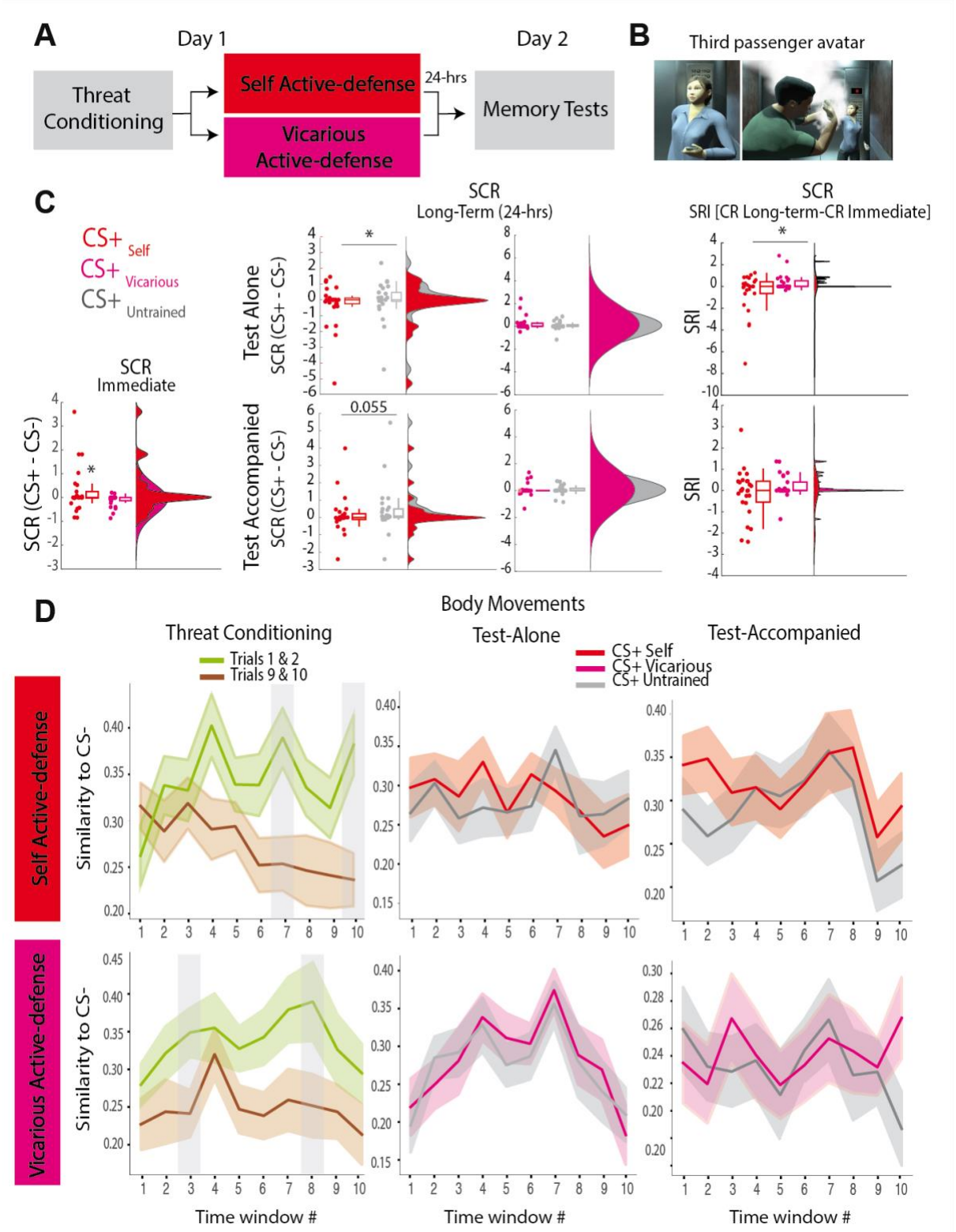
A
Day 1
Day 2
Threat Conditioning
Self Active-defense
Vicarious Active-defense
24-hrs
Memory Tests
B
Third passenger avatar
C
CS+ Self
CS+ Vicarious
CS+ Untrained
SCR Immediate
SCR (CS+ - CS-)
SCR Long-Term (24-hrs)
SRI [CR Long-term-CR Immediate]
Test Alone
Test Accompanied
SRI
0.055
D
Body Movements
Threat Conditioning
Test-Alone
Test-Accompanied
Trials 1 & 2
Trials 9 & 10
CS+ Self
CS+ Vicarious
CS+ Untrained
Self Active-defense
Vicarious Active-defense
Similarity to CS-
Time window #

Figure 4. Comparison of fear alleviation effects between Self and Vicarious ACT training. (A) The flow of Experiment 2 across two consecutive days. (B) Female passenger avatar defending against the violent avatar with a security spray on the participant's behalf in the Vicarious ACT training. (C) Immediate and long-term effects of Self and Vicarious ACT training on conditioned SCR (left and middle) and Spontaneous Recovery Index (SRI) (right). (D) Decreasing similarity in body movement patterns toward the CS+ versus the CS- avatars from early (trials 1-2) to late phase (trials 9-10) in both groups (left) and long-term effects of Self and Vicarious ACT training on body movements patterns (middle and right). Bayesian factor indicates anecdotal evidence for a greater similarity in body movement patterns between $CS+_{active}$ and CS- relative to the similarity between $CS+_{untrained}$ and CS- only following the Self ACT.

Methods

Experiment 1

Participants

Forty-one adults (20 males, Mean age = 25 ± 6 years old) without a reported history of anxiety or related disorders were paid to participate in the three-day experiments (8000 yen per day including agent charge). All participants provided written informed consent prior to their participation. A sample size of 39 was estimated based on our pilot testing and an a priori power calculation using G*power (F. Faul, Erdfelder, Lang, & Buchner, 2007) for detecting a conditioning effect in the VR setting with effect size of dz = 0.54 with a power (1-β err prob) of 0.95. The estimated sample size was corrected to 40 participants to ensure counterbalancing. Forty-one participants were enrolled in the end as one participant failed to complete the last session. One participant did not complete the reconditioning session but the data from other completed sessions were included in the analyses. This sample size is similar to a previous study with active avoidance among 28 human participants (Boeke et al., 2017).

Procedures

The Experiment consisted of four main experimental sessions conducted across three consecutive days: Conditioning (day 1), Extinction (EXT) and Active-defense (ACT) training sessions (day 2), and spontaneous secovery test (day 3). The test session on day 3 was followed by a renewal test session in a novel context and a reconditioning session. All sessions were conducted in virtual settings with a head-mounted VR display (HTC VIVE Pro Eye) (see Materials for details). The study was approved by an IRB committee at Sony (Tokyo)

Experimental sessions

Day 1 only consisted of a conditioning session with two CS+s and one CS-. In the conditioning session, three male avatars (Fig 1C) serving as CSs appeared on 10 trials each without showing any violence. The assignment of the three avatars to each CS was counterbalanced across participants. Two of the avatars appeared on additional 15 trials with violence, hence they both served as CS+s. One avatar never appeared with violence and thus served as a CS-. The task of participants was simply to observe the avatars throughout the session. On each trial, a participant was virtually situated in the back corner of an elevator. The trial sequence was as follows (Fig 1B): After 4 sec, the door of the elevator opened and one out of the three male avatar CSs entered the elevator after which the elevator door closed. In safe trials without violence, the avatar brunched around with his body facing toward the participant for 2 s and then turned around to face the closed door to remain standing semi-still for the rest of the trial (8 sec). On the trials with violence (CS+), an avatar entered the elevator and brunched with identical movements as in safe trials (CS-). However, after 6 s the avatar quickly walked toward the participant and raised his right arm to hit the participant's head. The experience of being hit in the realistic VR environment served as threatening unconditioned stimuli (US). The hitting action by the aviator was accompanied by a corresponding sound effect. The opening of the elevator door was also accompanied by a sound effect and a background machinery noise of the elevator was constantly played throughout each trial to enhance a realistic VR experience. Three avatars were all males

to reflect the relatively higher prevalence of violent incidence by males and related psychological traumas (Reichenberger et al., 2019).

The trial order was pseudorandomized with the restriction that no more than two consecutive trials of the same condition occur (e.g., a male avatar A with violence) to avoid habituation (Lonsdorf et al., 2017). Each of the initial two trials were always with one of the two CS+ avatars with violence in a counterbalanced order across participants. This was to boost expectation for aversive violence in the session (Bishop, Duncan, Brett, & Lawrence, 2004) and to avoid inducing any difference in potential *blocking* of conditioning effects (Meulders & Vlaeyen, 2012; Vaitl & Lipp, 1997) due to unequal numbers of safe trials with the two CS+s prior to the occurrence of initial violent trials.

On day 2, participants went through two training sessions: the Active-defense (ACT) and Extinction (EXT) sessions. The order of the two sessions was counterbalanced across participants and was separated by a 15 min break. Only one of the CS+s appeared during the ACT session ($CS+_{active}$) and only the other CS+ appeared during the EXT session ($CS+_{extinction}$), along with the CS- in both sessions.

In the ACT session, $CS+_{active}$ and CS- appeared 25 and 10 times, respectively. Among the trials with $CS+_{active}$, $CS+_{active}$ appeared 15 times showing the common entering and brunching actions followed by preparatory actions for hitting (i.e., approaching a participant to raise an arm over the participant) as in the first 9 sec of the CS+ trials in the Conditioning session (see Fig 1). However, here $CS+_{active}$ trials never ended with actual hitting due to preventative defense of spraying the $CS+_{active}$ avatar by participants. In the remaining 10 trials, $CS+_{active}$ showed no sign of violence as in the trials with CS-. Those trials without violence were included to equate the CS-US contingency rate between the Conditioning session on day 1 and the ACT sessions on day 2. Prior to the Active-defense session, participants were instructed to pull a trigger on a controller in their right hands to use a virtual defense spray against the $CS+_{active}$ avatar upon seeing his preparatory actions for hitting (i.e., approaching and raising a hand). The spraying action was practiced for one minute during the mirror session prior to the ACT session (see *Mirror session*). Upon pulling the trigger, the white mist was sprayed toward the $CS+_{active}$ avatar who displayed suffering in response to the released spray with his hands covering his face. Thus, only the CS+ but not the US (hitting) appeared during the active control session.

In the EXT session, the $CS+_{extinction}$ avatar appeared 25 times with the same action as CS- avatar (i.e., no violence) ($CS+_{extinction}$) along with CS- appearing 10 times. Thus, the total number of trials (35) was identical to that of the ACT session. In the EXT session, participants were simply instructed to observe the avatars.

The order of the trials in the ACT and EXT sessions were pseudo-randomized while avoiding three consecutive trials with the same condition. The context of the two training sessions was the identical elevator space as in the conditioning session. In the ACT session, the first two trials were always without precursor actions for violence to avoid the overestimation of the violence frequency in the upcoming trials (Bishop et al., 2004) relative to the EXT session.

Day 3 consisted of two test sessions and one reconditioning session. The first test, referred to as spontaneous recovery test, was performed in the original context as the conditioning session on day 1 and the training sessions on day 2. Next, to examine the effect of context on

fear memories (Vervliet, Baeyens, Van den Bergh, & Hermans, 2013)[48], a Renewal test was performed in a novel context. The novel context consisted of the same elevator space but with the light drastically dimmed. In both test sessions, the three CS avatars appeared eight times each without violence (US). The trial order was pseudorandomized with the first trial always being CS- to capture an irrelevant orienting effect (Boeke et al., 2017; Koizumi et al., 2016) and its data was not included in analysis. Finally, participants underwent a reconditioning session to investigate whether the training sessions were effective in reducing the susceptibility to repeated conditioning (X. An, Zhang, Liu, Yang, & Yu, 2019; Thompson, McEvoy, & Lipp, 2018). In the reconditioning session, the CS- avatar appeared eight times without violence, whereas the two CS+ avatars ($CS+_{active}$, $CS+_{extinction}$) appeared eight times without and eight times with violence. The context of the reconditioning session matched the original context. The first two trials of the reconditioning session were always with the two CS+s with violence in a counterbalanced order across participants to resemble the original conditioning session.

Mirror sessions

Prior to each experimental session, a participant was virtually situated in the back corner of an elevator for one minute to enhance immersive feeling (Slater et al., 2019). The elevator was identical to the experimental sessions except that a mirror appeared toward the wall on the left side from the participant. A mirror reflected him or herself as a human-like avatar. The avatar-self was semi-personalized to match the gender of each participant. A participant was asked to move around his or her body to observe the corresponding movements of the avatar-self. Only prior to the ACT session, the avatar-self appeared with a spray in the right hand and the participant practiced using a hand-held controller to trigger a virtual defensive spray to exert active defense movements.

Rating sessions

Participants provided subjective ratings on their fearful feelings and US (violence) anticipation upon seeing the CS avatars in the back corner of a virtual elevator. Throughout the experiment, they provided ratings for 6 times in total; before and after the conditioning session, after the Active avoidance, after the Extinction sessions, after the renewal test, and after the reconditioning session. In each rating session, the three CS avatars appeared once in a counterbalanced order across participants. Each avatar entered the elevator and paused with the panel appearing in front of him to display a question “how fearful do you feel upon seeing the man?”, followed by another question “how likely do you think the man is going to hit you?” in this order. Participants provided verbal reports of their rating on a seven-point scale, and an experimenter recorded the ratings.

Materials and equipments

The interactive VR images were programmed with the Unity 2019.3.0 and were presented at 60 Hz with a head mounted display, the HTC VIVE Pro Eye (2019, https://www.vive.com/eu/product/vive-pro-eye/overview/). The tracking of head positions (x, y, z) were also enabled by the HTC VIVE Pro Eye and the tracking of hand positions (x,y,z) were obtained by the Vive Trackers bound to the right hand. The image presentation and the recording of the tracked data were controlled by a laptop (G-Gear: N1588J-710/T) with NVIDIA® GeForce

RTX™ 2070, Intel® Core™ i7-9750H processor, and Windows 10 Pro. Eye-tracking was calibrated with the Super Reality (SR) runtime implemented in the HTC VIVE Pro Eye prior to each experimental session. Their skin conductance was measured by Shimmer3 GSR (LoAndStream v0.11.0, Shimmer®) with its electrodes attached to the participants' left palm. The skin conductance signal was recorded at 250 Hz with a Tobii Pro Lab software (ver 1.145.28180) installed in another laptop (Dell: Precision 7720) with Windows 10 Pro.

Body Movement

Positional and rotational data was obtained by the Vive Trackers and both head and hand motions of the subject in the VR were continuously tracked at 60 Hz during all experimental sessions. We obtained head and hand positions and rotations across the three axes (x, y and z), see Fig 1D. The sensor measurements of the trackers are provided in a head-fixed coordinate system where the x-axis points to the right, the y-axis points upward and the z-axis points forward.

Analysis

Subjective rating

Subjective ratings on fearful feelings and US anticipation were analyzed separately. Differential subjective ratings for both CS+s were calculated relative to the CS- within the same rating sessions. Exceptionally for the two training sessions of Experiment 1 on day 2, subjective ratings were analyzed only for the CS+ and CS- which appeared during the training session immediately before the rating session. This was to avoid including the rating data for the untrained CS+ due to the counterbalanced order of the two training sessions across participants.

SCR

SCR data were low and high pass filtered (0.015 - 0.05 Hz) with a FIR filter, detrended and normalized. The SCR onsets and peaks were automatically detected by Tobii Pro Lab software (Tobii Pro AB (2014). Tobii Pro Lab [Computer software]. Danderyd, Sweden: Tobii Pro AB.). The cumulative SCR was computed as the summation of all the through to peak amplitude deflections of individual SCR(s) within each trial, to effectively quantify the responses during the relatively long trial (14s) with a naturalistic VR scene with dynamically changing CS avatars. Only when the peaks and corresponding onsets of individual SCR(s) were observed within a trial, their values were accumulated as they may otherwise reflect irrelevant responses such as anticipation for the start or end of a trial. The cumulative SCR was log-transformed and scaled to the maximum cumulative SCR of the CS trials (not followed by US) of the conditioning session. For the conditioning and reconditioning sessions, the CS trials with US were not included in the analysis to exclusively examine the conditioned responses. The test sessions always started with one CS-presentation to capture the orienting response and was not included in analysis.

Movement Representational Similarity Analysis

To investigate the relationship between the patterns of participants' body movement and learned threat associations we used Representational similarity analyses (RSA) (Kriegeskorte et al., 2008; Lu & Ku, 2020; Visser et al., 2015). We calculated the similarity of body movements upon seeing different CS avatars based on the position and rotation of both head and hand for the three axes

(12 dimensions, see Fig 1D). To obtain a measure that is independent from coordinate values and unbiased by previous positions during a trial, we calculated the absolute difference between two consecutive recorded samples (|Position(t2)-Position(t1)|) within trials. This yielded a measure of sample-by-sample change in movement for each of the 12 movement parameters. For all analyses, CS+ trials with a US (violence) were discarded, such that remaining CS+ trials were identical to the CS- trials except for the avatar identity ($CS+_{active}$,$CS+_{extinction}$, or CS-, see Fig 1C). Next, to reduce the amount of comparisons we divided the data within a trial into 10 time windows of 833 ms (50 samples) each. We created a vector per condition (e.g., CS+ or CS-) for each participant per time window (Lu & Ku, 2020) (https://neurora.github.io/NeuroRA). Each vector now contained the spatial pattern of movement (12 movement parameters) within a specific time window (50 time samples), related to a particular condition. Pairwise Pearson correlations were calculated between vectors of condition pairs, resulting in a similarity measure, depicting correlations among conditions for each participant per time window. For instance, the correlation between movement in response to the CS+ and CS- avatar provided for each time window of the trial, and for each participant a metric of how similar the movement in (non-violent) CS+ trials was relative to CS- trials. Correlation coefficients were then Fisher-transformed and averaged across participants. Finally, we used classical and bayesian paired t-tests to compare correlation coefficients.

The RSA was first conducted to assess threat conditioning by comparing the CS+ to CS- similarity between the early and the late phases of the conditioning session (trials 1-2 versus trials 9-10, one-sided paired t-tests). We expected the CS+ to CS- similarity to decrease in the late phase of the conditioning session due to emergence of threat conditioning effects in body movements. Second, RSA was performed to compare the effect of active defense and extinction training on spontaneous recovery on day 3 by comparing $CS+_{active}$/CS- similarity versus $CS+_{extinction}$/CS- similarity in the first two trials of the test session. We expected the ACT training to be more effective in reducing the fear response, resulting in greater $CS+_{active}$/CS- similarity compared to $CS+_{extinction}$/CS- similarity.

Statistical Analysis

Where Shapiro-Wilk's test indicated that normality assumptions were violated, Wilcoxon correction was applied. In addition to traditional null hypothesis significance testing (NHST), we computed Bayes factors (BF) for all analyses. BF allowed us not only to find evidence for our tested hypotheses but to quantify the evidence in favor of the null hypothesis. We based our results on BF especially when it was beneficial to infer the relative degree of evidence across multiple time windows (10) within a trial to capture the temporal dynamics in the body movement analyses. For the Bayesian analyses a Cauchy prior distribution centered around zero was used with an interquartile range of $r = 0.707$. All analyses were performed using JASP (Version 0.16). For the Bayesian analysis in JASP a Cauchy prior distribution centered around zero was used with an interquartile range of $r = 0.707$.

When addressing the main directional hypotheses, one-tailed t-tests and directional Bayesian paired samples t-tests (BF_{+0}) were used, otherwise two-tailed t-tests and non-directional Bayesian paired samples t-tests (BF_{10}). To test for evidence supporting the null hypothesis, we performed Bayesian paired samples t-tests (BF_{01}). We used a common interpretation of the Bayes factors

(BF 1-3: anecdotal evidence, BF 3-10: moderate evidence, BF>10: strong evidence) (Ly, Verhagen, & Wagenmakers, 2016). For the classical t-tests, we corrected for multiple comparisons by fixing the false discovery rate (FDR) at 0.05 (Benjamini & Hochberg, 1995). First, we used one-tailed t tests in validating the conditioned effect (i.e., larger responses to CS+ than to CS-). Second, we used one-tailed t-tests to examine the more robust fear reduction effect of the ACT training than the EXT training when tested 24 hrs after the training (i.e., long-term effect), as predicted based on previous literature (Boeke et al., 2017). The changes of SCR and movement data within or across sessions were analyzed by examining responses in the early and late phases (two trials each) within a session. As the subjective ratings were collected only once after each session type, the data were not divided into smaller phases. To examine the effects of participants' gender and training order on fear attenuation in Experiment 1 we used Repeated Measures ANOVA and Tukey test for multiple comparison.

We did not exclude any participants based on the presence of differential SCR to CS+ relative to CS- during the conditioning. This is because a recent perspective points that those who do not show differential responses are generally more prone to anxiety and we may lose a more relevant cohort to the topic of fear memories (Beckers et al., 2013). In addition, we did not use SCR as an exclusion criteria given that we aimed to evaluate the degree of threat-conditioning multidimensionally, not only with SCR but with additional measures of bodily movements and subjective ratings. However, we shared how results remain quantitatively similar after removing SCR-based outliers across sessions.

Experiment 2

The methods for Experiment 2 were generally similar to Experiment 1 with a few exceptions summarized in the following sections.

Participants

Fifty-four adults (27 males, Mean age = 23 ± 3 years old) were paid to participate in the two-day Experiment (1050 yen per hour). Participants had no issue with their athletic abilities and visual acuity and all were right-handed. Participants were divided into two groups (Self-ACT group: n = 27, 14 males, Mean age = 24 ± 2; Vicarious-ACT group: n = 27, 13 males, Mean age: 24 ± 3) One additional participant from the Self-ACT group stopped the Experiment due to a malfunction of the equipment and the data was not included.

Procedures

Experiment 2 was conducted across two consecutive days and only Active-defense (ACT) training sessions (day 1) and spontaneous recovery tests (day 2) differed from Experiment 1.

Active defense training sessions:

In the ACT training session, participants in the Self ACT group underwent the same procedure as in Experiment 1, where they actively defended themselves with a virtual security spray to prevent violence. Participants in the Vicarious ACT group watched a third passenger avatar defending against the violent avatar with a security spray on their behalf. In both groups, the third avatar

accompanied the participants in the training sessions to equate the potential effect of social presence in fear responses (Eisenberger et al., 2011) with the Vicarious ACT group. The third avatar was a female passenger that stood in front of the participant's standing position, facing the center of the elevator, in front of the number button on the right side of the elevator door (Fig 4A). The gender of the third avatar was fixed to be female across participants to make her clearly distinguishable from the CS male avatars so that participants would not potentially anticipate threat (violence) from the third avatar. When the male avatar CS was going to attack in the ACT training session, the third avatar sprayed the virtual security spray toward the head of the avatar at the timing when the paired self-group participant sprayed on the same trial. When the avatar CS did not inflict violence, the third avatar operated her smartphone to reflect a common encounterance scene with others in an elevator. The third avatar, however, lowered her smartphone and looked up at the male CS avatar when the participants in the Self-ACT group initiated the spraying action, as absence of reaction from the third avatar would otherwise appear unrealistic. Since subjects did not undergo an EXT training, one CS+ was associated with ACT training ($CS+_{active}$) and the other CS+ was untrained ($CS+_{untrained}$).

Spontaneous recovery tests:

Day 2 consisted of two spontaneous recovery test sessions and one reconditioning session. The first test (Test-Alone) was performed in the exact same context as in the conditioning session and as in Experiment 1. The second test (Test-Accompanied) was performed in the same context as in the conditioning session but adding a third avatar in the scene to examine the effect of social presence on fear memories retrieval. Participants were informed that a passenger would appear prior to the Test-Accompanied.

Rating sessions

*P*articipants provided subjective ratings for 6 times in total; before and after the conditioning session, after the Active-defense sessions, after both the first and second spontaneous recovery tests and after the reconditioning session.

Statistical Analysis

Statistical methods were identical to those in Experiment 1. In addition, to examine the effects of ACT group on fear alleviation effects, we used Repeated Measures ANOVA and Tukey test for multiple comparison. We used independent t tests to examine differences of CR between groups.

Discussion

In the present study, we demonstrated that human participants develop threat-conditioned body movement patterns to discern fear versus safety cues in a naturalistic virtual setting. Furthermore, the subsequent training to actively defend against the threat with the participants' own body movement led to the weakening of the conditioned body movement patterns as well as a long-term (24 hrs) reduction of the physiological responses and subjective fear toward the threat-conditioned cues. While the Extinction (EXT) training led to better immediate fear reduction effect than the Active-defense (ACT) training, the advantage of the EXT training was transient. Unlike the EXT training, the ACT training resisted to spontaneous recovery when tested 24 hrs later. These results highlight the critical role of body movements in both acquisition and reduction of fear among humans. Through experience humans can opt their multidimensional body movements in anticipation of the cued threats to avoid or minimize harm, and they can alleviate their fear state by training an alternative bodily defense movements.

Fear acquisition in bodily action

The results that body movement patterns are differentiated through threat association learning are in keeping with previous animal literature. The non-human animal literature often postulates their body movements as one primary consequence of acquiring fear memories (Bouton & Bolles, 1980; Gruene, Flick, Stefano, Shea, & Shansky, 2015). Our study extends the literature to show that humans are not exceptions. To our knowledge, our study provides the first direct evidence that body movement signals threat conditioning in humans. In human studies, their body movements are generally marginalized compared with other common measures such as physiological, subjective, or neural responses. Although body movements may appear evolutionary ancient and thus less relevant to humans, various modern-age threats, such as interpersonal violence and fire, still require one's whole-body movement for protection. Thus, threat-conditioning should enhance survival if such learning results not only in physiological, perceptual or cognitive changes but also in actual body movements to directly defend oneself against the threat.

Fear alleviation through bodily defense

Our results further show that physical training to actively defend oneself against a threat has fear alleviation effects. Such results were robust in that they were consistently observed across physiological response, subjective experience, and to some extent (as indicated as moderate evidence with Bayesian factor) body movement patterns *per se*.

Crucially, the Active-defense training was resistant to spontaneous recovery, one of the major constraints of conventional extinction training. Generally, fear alleviation effect of extinction training is transient and reemerges with the passage of time in the absence of any further training (Myers & Davis, 2007; Robbins, 1990). Here we replicated the immediate fear alleviation effects with our virtual Extinction training, followed by the spontaneous recovery of fear responses 24 hrs later. This result supports the theoretical and experimental studies suggesting that extinction is a form of inhibitory learning rather than an erasure of acquired fear (B. An et al., 2017; Myers & Davis, 2007).

In contrast, the effect of active-defense training was absent when assessed immediately on the day of the training but the training led to a long term effect to prevent spontaneous recovery. The absence of spontaneous recovery with the active-defense training suggests that its underlying mechanism may not rely on the inhibitory mechanisms as the conventional extinction training. Supporting this possibility, our preliminary results showed that a larger immediate fear reduction effect through the extinction training was correlated with a larger spontaneous recovery on a subsequent day across participants, suggesting that suppression of fear response during the extinction training may be uninhibited with a passage of time (see Supplementary results of Experiment 1 - SCR section). Meanwhile, the immediate fear reduction effect through active-defense training showed no such correlation with the next day recovery of fear response.

The underlying neural mechanisms for the fear alleviation effect of the embodied active-defense training remains to be studied in future studies. Previous studies with human participants have elucidated the role of network among medial Prefrontal Cortex (mPFC), amygdala, and striatum in successful performance of active avoidance (Collins et al., 2014; Wanke & Schwabe, 2020). Furthermore, studies with rats suggest a more detailed mechanism in which mPFC mediates the trade-off between passive fear-like responses and active avoidance (Ledoux & Daw, 2018; LeDoux et al., 2017; Martinez et al., 2013). Future studies may examine whether the training of bodily active defense as in the current study likewise reduces the passive fear responses via mPFC. Such studies may further examine whether and how the embodied active avoidance procedure as in our study may recruit additional mechanisms to alleviate fear.

The current study extends the previous human studies with active-avoidance by directly demonstrating the efficacy of embodied defense movements in alleviating fear. As described in the Introduction, the animal active-avoidance training has been translated to human studies (Boeke et al., 2017). Yet, generic technical limitations of human behavioural studies only allowed an ecologically less valid form of active avoidance, such as moving an icon on a monitor through a key pressing by stationalized participants (Boeke et al., 2017; Wanke & Schwabe, 2020). While those studies successfully demonstrated the potential fear alleviation effects of active avoidance training , such a less embodied method may limit its applicability to real-life threats - such as interpersonal violence, fire in a house, and immencing vehicles - against which humans need to use their own bodies to defend and minimize actual harm.

By leveraging on the recent progress of VR technologies, we unleashed the naturalistic defensive body movements of the participants. The demonstrated fear alleviation effects of such an embodied active defense training should extend the applicability of the fear alleviation training for various types of human fear through defense. Recently, the extinction of fear in VR settings has acquired popularity in clinical settings. Yet, a meta-analysis holds that extinction with VR may be only slightly more effective than non-VR extinction procedures (Powers & Emmelkamp, 2008). Here, our results show the promising approach to better leverage the VR techniques to improve treatments for anxiety and fear-related disorders such as PTSD. Future studies may further extend its applicability by directly testing the effects of active defense with a wider varieties of common traumatic scenarios.

Despite the aforementioned advantage of the ACT training over the EXT training, some of the limitations should be cautiously considered when potentially applying the ACT training for fear alleviation. First, the ACT training led to higher SCR to CS+ than the EXT training when the immediate effects of the training were examined (Fig 3B). Second, the ACT training generally enhanced US anticipation (Fig 3A) potentially because of the increased number of trials where CS+ was further reinforced with precursors of US (i.e., gesture for violence without resulting in hitting). Last and relatedly, fear was more likely to be reconditioned after the ACT training although such an effect was only transient (Supplementary). These results suggest that some individuals might momentarily experience more distress with the ACT training than with the EXT training, even though the ACT training would result in a better long-term fear alleviation effect. Thus, the ACT training may be applicable to reduce fear of individuals when they could readily tolerate those limitations to achieve better long-term fear alleviation effects.

Crucially, our results from Experiment 2 revealed that participants' own bodily defense action is necessary for the fear alleviation effects of Active-defense. Specifically, the Active-defense training led to a long-term fear reduction only when participants executed the defense action (spraying), while mere observation of a third avatar defending on their behalf did not. This finding is in contrast with previous studies that both fear acquisition and extinction can take place through observation (Golkar, Haaker, Selbing, & Olsson, 2016; Golkar, Selbing, Flygare, Öhman, & Olsson, 2013; Haaker et al., 2021; Keum & Shin, 2019). Unlike acquisition and extinction of fear, our results suggest that fear alleviation through active defense does require participants' own action and vicarious defense is not sufficient to reduce fear.

The fear alleviation effect of the participants' own defense movement could be due to their body movement, sense of control, or both. Our preliminary results showed that duration and reaction speed of participants' defense movements correlated with success in fear reduction as assessed with body movement pattern similarity (Fig S1C). Although the results were suggestive, they suggest at least some contribution of the movement itself which cannot be explained away by the sense of control. As sense of control is intrinsic to bodily defense movements, the contributions of movement and sense of control may be studied more systematically in the future.

Interestingly, we found that male participants were more likely to benefit from the active-defense training relative to the extinction training (Fig S4B). However, the results seem to be partly due to the fact that male participants were more susceptible than female participants to show spontaneous recovery after the Extinction training. As the use of male violent avatars limits our interpretation of the participants' gender effect, future studies may more directly pursue the gender effect.

The current finding elucidates the potential of multivariate body movement patterns as a novel measure of fear states in humans. The two dominant measures of human fear states, physiological responses and subjective experience, appear to rely on distinct neural underpinnings (Taschereau-Dumouchel, Kawato, & Lau, 2019). The addition of body movements as a third measure in future studies may extend the quest of multidimensional human fear circuits working in concert.

In summary, we demonstrate the critical role of body movements in both acquisition and alleviation of human fear memory. The findings may benefit the field by calling to re-centralize the body in research of human fear and its disorders such as PTSD.

Supplementary

Supplementary results of Experiment 1

SCR

During the conditioning session, participants (n=41) displayed numerically higher SCR toward both $CS+_{active}$ (M = 0.242 ± s.e. 0.021) and $CS+_{extinction}$ (0.239 ± s.e. 0.022) relative to CS- (0.225 ± s.e. 0.022) but the difference from CS- (i.e., Conditioned Response, CR) was non-significant [one-tailed CS+ > CS-; $CS+_{active}$ $t(40) = 0.91$, $p = 0.18$; $BF_{+0} = 0.403$; $CS+_{extinction}$ $t(40) = 0.71$, $p = 0.23$; $BF_{+0} = 0.32$] (Fig S1A). Lack of CR on day 1 may be because of the generally elevated SCR including the CS- trials during the conditioning session where frequent violence was experienced and thus was anticipated.

With Spontaneous Recovery Index (SRI) [change of CR from the late training on day 2 to the early phase test on day 3], $CS+_{extinction}$ showed a trend of positive SRI while $CS+_{active}$ did not [CR early phase test day 3 Vs CR late phase training day 2; $CS+_{extinction}$ $t(40) = 1.97$, $p = 0.055$; $BF_{+0} = 0.98$; $CS+_{active}$ $t(40) = 0.59$, $p = 0.55$; $BF_{+0} = 0.19$] (Fig 2B). Interestingly, participants that showed a lower CR to the $CS+_{extinction}$ in the late phase of training (last two trials) on day 2 showed a larger CR to the $CS+_{extinction}$ in the early phase (first two trials) of the Test on day 3 [Pearson correlation R = -0.58, $p < 0.001$; $BF_{+0} = 529.8$]. Given that extinction training is speculated to largely depend on the suppression of the originally acquired conditioning effect [66], this result suggests that participants with larger suppression of the CR to the $CS+_{extinction}$ during the EXT training are more prone to the return of CR with a passage of time (24 hrs). CR to the $CS+_{active}$ in the late training on day 2 did not correlate with CR to the $CS+_{active}$ in the early phase of test on day 3 [R = 0.12, $p = 0.42$; $BF_{+0} = 0.26$; r ACT Vs r EXT z = -3.41, $p = 0.002$], suggesting a potentially different mechanism of fear reduction with the ACT training than suppression with conventional extinction training.

We finally examined whether ACT and EXT training had an attenuation effect on CR when participants were reexposed to the CS+s in a novel context test and reconditioned on day 3. Participants did not show any CR to the $CS+_{extinction}$ and the $CS+_{active}$ in both the early or late phases of the Renewal Test, and CR did not differ between the two CS+s [$CS+_{extinction}$ Vs $CS+_{active}$; early phase, $t(40) = 0.15$, $p = 0.87$; $BF_{+0} = 0.17$; late phase, $t(40) = 0.74$, $p = 0.46$ $BF_{+0} = 0.21$]. This suggests that unreinforced trials in the spontaneous recovery test in the original context extinguished CR to both CS+s before the novel-context test. In the reconditioning session participants showed statistically similar levels of SCR to the $CS+_{extinction}$ and $CS+_{active}$ [$t(40) = 0.81$, $p = 0.41$; $BF_{+0} = 0.23$]. However, an analysis with the early phase of the training revealed that they showed a significantly smaller SCR to the $CS+_{extinction}$ relative to CS- [$t(40) = -3.11$, $p = 0.003$; $BF_{+0} = 10.34$] while there was no significant difference between $CS+_{active}$ and CS- (i.e., CR) [$t(40) = -0-25$, $p = 0.801$; $BF_{+0} = 0.17$]. This suggests that the EXT training protected them from getting reconditioning in the early phase by suppressing responses to the trained CS+ (Fig S1A). Yet, such effect was transient as participants did not show any CR to neither the $CS+_{extinction}$ nor to the $CS+_{active}$ in the late phase of reconditioning [$CS+_{extinction}$; $t(40) = 1.04$, $p = 0.305$; $BF_{+0} = 0.7$;

$CS+_{active}$; $t(40) = 1.45$, $p = 0.15$; $BF_{+0} = 0.44$]. The CR did not differ between the CS+s in the late phase of reconditioning [$t(40) = 1.108$, $p = 0.27$; $BF_{+0} = 0.29$] (Fig S1A).

Additionally, we examined whether the gender of participants had any effect on ACT and EXT training effects. Threat conditioning was unaffected by gender but we found a significant interaction between gender and training in the long-term effect [n females = 21; n males = 20; $F(1,39) = 4.76$ $p = 0.03$] and in the SRI [$F(1,39) = 5.94$ $p = 0.01$], along with a significant main effect of the training in the SRI [$F(1,39) = 7.79$ $p = 0.008$] . Only males showed a more robust long-term reduction of SCR to $CS+_{active}$ relative to $CS+_{extinction}$ [$t(19) = -2.51$, $p = 0.02$; $BF_{+0} = 2.77$] and a higher SRI to $CS+_{extinction}$ than to $CS+_{active}$ [$t(19) = -3.27$, $p = 0.004$; $BF_{+0} = 11.18$; multiple comparisons $t(19) = -3.61$, $p\ tukey = 0.004$]. These results suggest that males may benefit more from the ACT training as indicated by their higher reemergence of CR 24-hrs after the EXT training while a stronger reduction of CR 24-hrs after the ACT training (Fig S4B).

We next examined whether the order of the training sessions - whether the participants first underwent the ACT or EXT training session - had any influence on fear attenuation effects. We found no interaction between the order of the training sessions and the CS type ($CS+_{active}$/$CS+_{extinction}$) in both immediate and long-term effects [1st AA N = 20; 1st EXT N = 21; Immediate: $F(1,39) = 0.17$ $p = 0.67$; Long-term: $F(1,39) = 0.006$ $p = 0.93$] (Fig S4C).

Subjective ratings

After the conditioning session on day 1 participants reported higher feelings of fear as well as higher anticipation of violence toward both $CS+_{active}$ and $CS+_{extinction}$ relative to CS-. Yet, subjective ratings of fear were higher for $CS+_{extinction}$ than $CS+_{active}$ [two tailed; $CS+_{extinction}$ Vs $CS+_{active}$ $t(40) = -2.10$, $p = 0.042$; $BF_{+0} = 0.81$] while subjective rating of US-likelihood were statistically similar between the CS+s [two tailed; $CS+_{extinction}$ Vs $CS+_{active}$ $t(40) = -0.43$, $p = 0.66$; $BF_{+0} = 5.42$]. Note that this initial difference in fear ratings unlikely contributed to the critical long-term fear reduction effects, as the initial difference (larger fear rating to $CS+_{active}$) was in the opposite direction from the long-term effect (generally reduced fear responses to $CS+_{active}$).

We finally examined whether the ACT and EXT training reduced subjective rating of fear and US-likelihood after the reconditioning session on day 3. Participants showed higher subjective ratings to CS+s than CS- after the reconditioning sessions [Fear: CS+ Vs CS-; $CS+_{active}$; $t(34) = 5.67$, $p < 0.001$; $BF_{+0} > 100$; $CS+_{extinction}$; $t(34) = 3.96$, $p < 0.001$; $BF_{+0} > 100$; US-likelihood: $CS+_{active}$; $t(34) = 8.07$, $p < 0.001$; $BF_{+0} > 100$; $CS+_{extinction}$; $t(34) = 5.26$, $p < 0.001$; $BF_{+0} > 100$]. Subjective ratings of US-likelihood, but not subjective ratings of fear, were higher to $CS+_{active}$ than to $CS+_{extinction}$ after Reconditioning [$CS+_{active}$ Vs $CS+_{extinction}$; US-likelihood $t(34) = 2.25$, $p = 0.03$; $BF_{+0} = 1.68$; fear $t(34) = 1.04$, $p = 0.30$; $BF_{+0} = 0.30$]. These results suggest that ACT training led to a higher subjective anticipation of violence (US) after the reconditioning session (Fig S1B), which is expected from the additional experience of US-reinforced $CS+_{active}$ trials during the ACT training session. Despite this, ACT and EXT training had a comparable effect on subjective rating of fear.

Correlations between body movements and SCR measures

Finally, since both SCR and body movements showed convergent long-term fear reduction effects of ACT defense on day 3, we examined whether the long-term fear reduction effects in physiological responses and those in body movements correlated with each other. We found a specific correlation after the ACT training [r = -0.37, *p* = 0.015] that was significantly different from the correlation found after EXT training [r = -0.01, *p* = 092; ACT Vs r EXT z = -1.65, *p* = 0.047] (Fig S1C). This suggests that after training ACT defense, those subjects that showed higher body movement similarity to safety (CS-) in response to $CS+_{active}$ also showed lower physiological CR.

Next, we examined whether the magnitude of the physical defense (spraying) action during the ACT training predicted the long-term fear alleviation effects on day 3. Specifically, we explored whether spraying behavior influenced the fear alleviation effects of ACT training in those time windows where ACT training outperformed EXT training in the early test (T6 and T7). In Experiment 1, we found that participants who showed longer defensive actions and slower defensive Reaction Times (RT) during the ACT training developed more body movement similarity between $CS+_{active}$ and safety (CS-) when tested 24hrs later [spray duration r = -0.31, *p* = 0.04; spray RT r = 0.33, *p* = 0.03]. These results suggest that defensive bodily action itself during the ACT training may contribute to the fear alleviation effects beyond the mere perception of the acquired controllability over the threat. However, the results should be treated as suggestive as they were found only in the time window #6 but not the other time window #7, the two windows in which moderate evidence for the difference in the body movement patterns for $CS+_{active}$ and $CS+_{extinction}$ was observed in the early Test (see Results of Experiment 1). Similarly, participants from Experiment 2 who underwent the Self ACT defense training, also showed that slower defense RTs predicted more similarity in the body movement patterns between CS+ and CS- in the Test-Accompanied [spray RT Self-ACT r = 0.57, *p* = 0.002]. However, these results are also suggestive as they were found only in the time window #10 but not the other time window #2, where ACT training outperformed no-training effects based on body movement patterns (see Results of Experiment 2). In contrast, participants who underwent the Vicarious-ACT training did not show any correlation between yoked spraying behavior and fear alleviation effects [r = -0.20, *p* = 0.33; Self-ACT Vs Vicarious-ACT z = 2.88 *p* = 0.002]. While it may be counterintuitive to find a correlation between longer RT (i.e., slower onset of defensive action) and attenuation of fear-like body movement patterns, one possibility is that participants with longer RT may have acquired higher confidence in their capability to defend against the approaching threat of CS+ avatar, leading to longer observation of the avatars before taking a defensive action. One alternative possibility could be that participants with lower fear in the late conditioning session (i.e., higher similarity in the body movement patterns between $CS+_{active}$ and CS-) may later show slower defense RT in the ACT training session. However, body movement patterns during the late conditioning session in those time windows where similarity between CS+s and CS- decreased relative to the early conditioning session (Experiment 1: T1 and T8-T10; Experiment 2: T7 and T10) did not correlate with spray behavior.

Supplementary results of Experiment 2

SCR

Repeated measures ANOVA with one between-subjects factor of training type (Self and Vicarious ACT training) and one within-subjects factor of CS+ type ($CS+_{active}$ and $CS+_{untrained}$) showed that there was neither a significant main effect of ACT training nor an interaction between ACT training type and CS+ type in the conditioning session [ACT training type: $F(1, 50) = 0.77$, $p = 0.38$; ACT training type * CS+type: $F(1, 50) = 1.56$, $p = 0.21$]. Repeated measures ANOVA on CR towards $CS+_{active}$ ($CS+_{active}$ - CS-) in the early ACT training session also showed no effect of ACT training nor interaction [ACT training type: $F(1, 50) = 0{,}50$, $p = 0.48$; ACT training type*CS type: $F(1, 50) = 3.48$, $p = 0.06$]. The main effect of CS+ was also non-significant [conditioning session: $F(1, 50) = 0.04$, $p = 0.82$; early ACT training session: $F(1, 50) = 2.91$, $p = 0.09$], indicating a statistically similar level of CR to both $CS+_{active}$ and $CS+_{untrained}$ prior to the training sessions. The CR to both CS+s ([CS+ active: M = 0.020 ± s.e. 0.017; $CS+_{untrained}$: M = 0.007 ± s.e. 0.014], however, were not significantly different from zero (i.e., not significantly larger than CS-) [$CS+_{active}$: $t(51) = 1.16$, $p = 0.12$; $BF_{+0} = 0.49$; CS+untrained: $t(51) = 0.48$, $p = 0.31$; $BF_{+0} = 0.23$). During the early ACT training session CR to $CS+_{active}$ [M = 0.067 ± s.e. 0.097] was also no significantly different from zero [$t(51) = 0.69$, $p = 0.24$; $BF_{+0} = 0.28$]. This absence of significant CR is in line with the day 1 results from Experiment 1 that CR from the current realistic conditioning emerges only on subsequent days (see Main text). However, given that the critical effects of the training were still observed clearly in Experiment 1, we went on to examine the planned effects in test sessions on day2 of Experiment 2.

Subjective ratings

After the ACT training participants from both groups showed an immediate reduction in subjective fear ratings, but not US anticipation, to the $CS+_{active}$ relative to the post-conditioning ratings on day 1 [Self ACT: Fear; $t(26) = 2.25$, $p = 0.03$; $BF_{10} = 1.74$, US-likelihood; $t(26) = -0.27$, $p = 0.78$; $BF_{10} = 0.21$; Vicarious ACT: Fear; $t(26) = 2.67$, $p = 0.01$; $BF_{10} = 3.82$, US-likelihood; $t(26) = 0.12$, $p = 0.90$; $BF_{10} = 0.20$]. These results suggest that both Self and Vicarious ACT training did not immediately reduce the US-expectancy while they did alleviate subjective fear, potentially due to the violent nature of the defense action. Meanwhile, both groups showed a long-term reduction of both fear and US anticipation ratings to the $CS+_{active}$ on day 2 relative to the post-conditioning ratings on day 1 [Test-Alone; Self ACT: Fear: $t(26) = 4.39$, $p < 0.001$; $BF_{10} > 100$, US-likelihood: $t(26) = 5.41$; $p < 0.001$; $BF_{10} > 100$; Vicarious ACT: Fear: $t(26) = 4.84$, $p < 0.001$; $BF_{10} > 100$, US-likelihood: $t(26) = 5.70$, $p < 0.001$; $BF_{10} > 100$] [Test-Accompanied; Self ACT Fear: $t(26) = 5.71$, $p < 0.001$; $BF_{10} > 100$, US-likelihood: $t(26) = 7.75$, $p < 0.001$; $BF_{10} > 100$; Vicarious ACT Fear: $t(26) = 4.40$, $p < 0.001$; $BF_{10} > 100$, US-likelihood: $t(26) = 5.12$, $p < 0.001$; $BF_{10} > 100$] (Fig S5).

Next we assessed the differences in long-term subjective ratings following the Self and Vicarious ACT training. For each test, we conducted a repeated measures ANOVA with one between-subjects factor of training type (Self and Vicarious ACT training) and one within-subjects factor of CS+ type ($CS+_{active}$ and $CS+_{untrained}$). In the Test-Alone we found no interaction between training type and CS+ type for the subjective ratings of both fear and US-likelihood [$F(1, 52) = 0.02$, $p =$

0.88; F(1, 52) = 1.28, p = 0.26, respectively]. However, in the Test-Accompanied we found a significant interaction with the US-likelihood ratings [F(1, 52) = 5.47, p = 0.02]. Multiple comparisons revealed that the Vicarious ACT group showed higher US-likelihood ratings toward $CS+_{active}$ than $CS+_{untrained}$ [$t(26) = 3.56$, $p\ tukey = 0.005$] when the Self ACT group did not [$t(26) = 0.25$, $p\ tukey = 0.99$] (Fig S5). With the fear ratings, we found no significant interaction between training type and CS+ type in the Test-Accompanied [F(1, 52) = 2.50, p = 0.12]. This result however should be treated as suggestive as we observed a similar trend of higher subjective fear rating for $CS+_{active}$ than $CS+_{untrained}$ in the Vicarious ACT group already in the post-conditioning session. However, the result remained quantitatively similar after removing outliers with unbalanced ratings in the post-conditioning. These results suggest that the social presence of the third avatar, who was present during the training but not during the conditioning, might facilitate violence anticipation especially when the third avatar, rather than the participants themselves, played a central role in regulating threat situations as in the Vicarious ACT group.

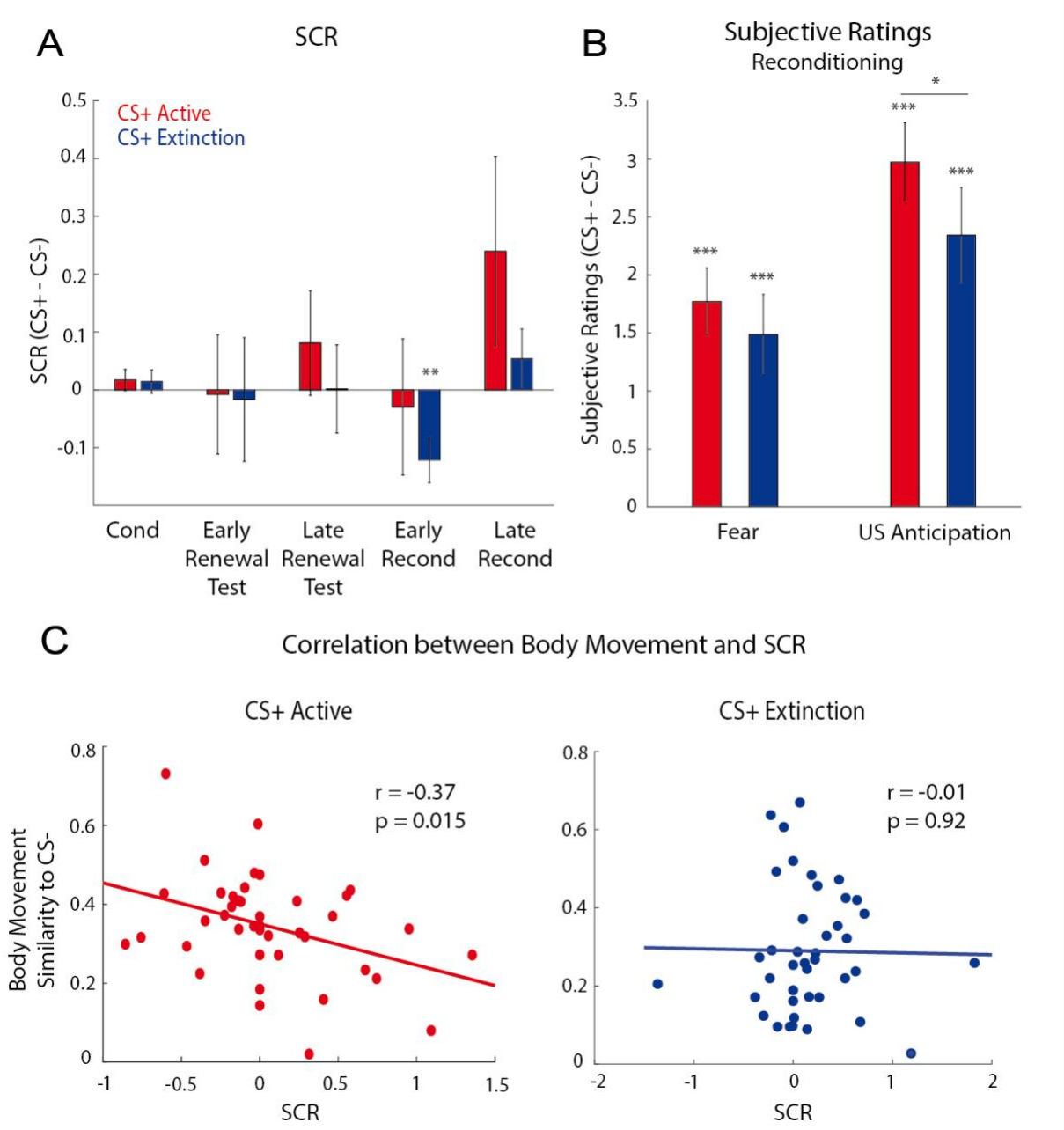

Figure S1. (A) Bar graphs representing conditioned responses (CR) [CS+ - CS-] based on SCRs during the conditioning session on day 1, during the early and late phases of the Renewal Test and during the early and late phases of Reconditioning session on day 3, all from Experiment 1. (B) Bar graphs representing conditioned responses (CR) [CS+ - CS-] based on subjective fear ratings after the Reconditioning phase on day 3 of Experiment 1. (C) Scatter-plots with least-square regression lines of the intrasubject correlations between SCRs and body movements after 24-hrs of training in the Spontaneous Recovery Test. Error bars represent the SEM. P-values are based on two-tailed paired t-tests. ACT= Active Avoidance Training; EXT = Extinction Training (related to Figure 2B and 3B).

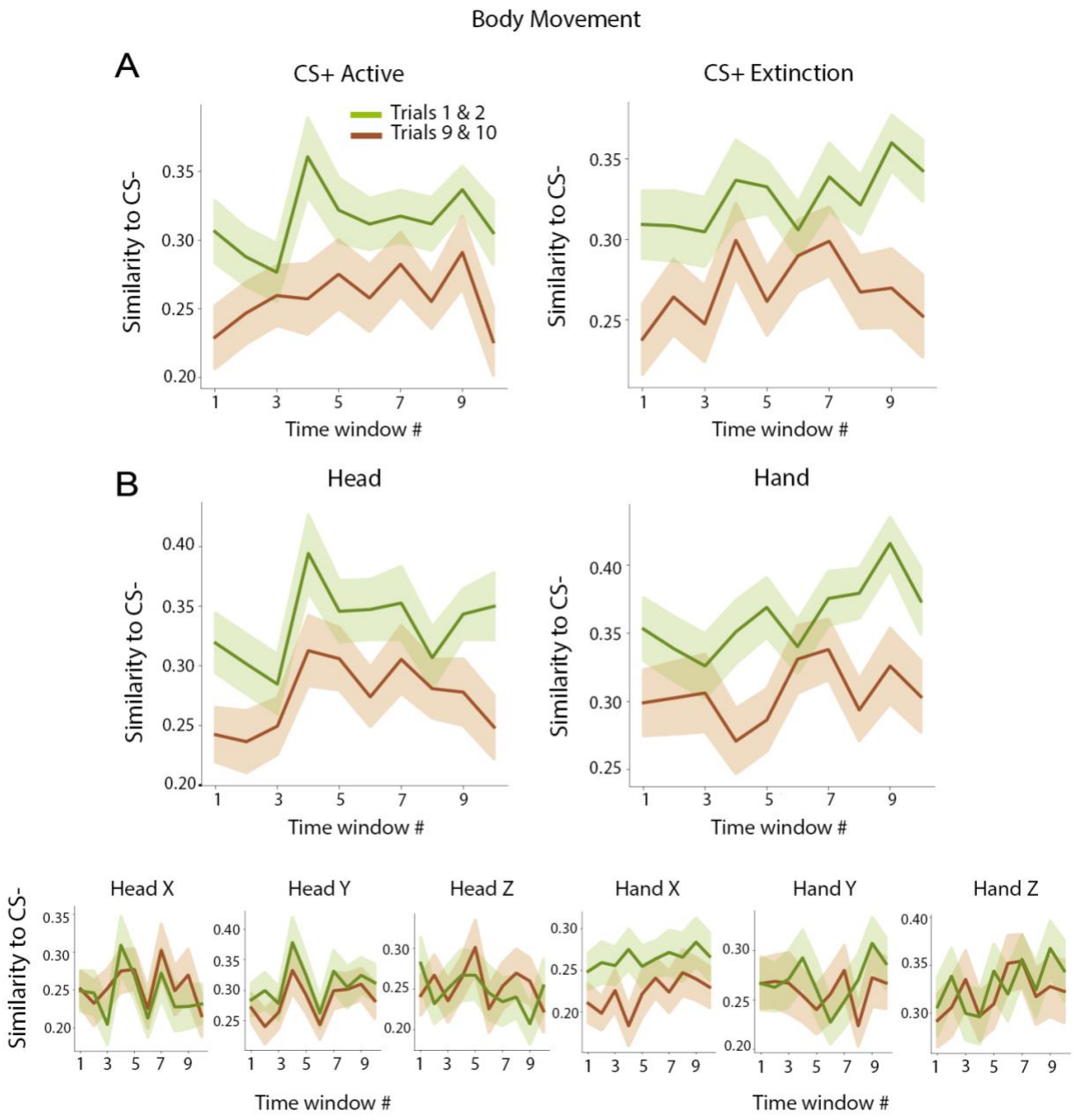

Figure S2. (A) Decreasing similarity in body movement patterns between each CS+ avatar (CS+ Active and CS+ Extinction) versus the CS- avatar from early (trials 1-2) to late phase (trials 9-10) of the conditioning session from Experiment 1. (B) Upper panels: Decreasing similarity in both head and hand movement patterns toward CS+ avatar versus the CS- avatar from early (trials 1-2) to late phase (trials 9-10). Lower panels: Body movement patterns toward CS+ avatar versus the CS- avatar from early (trials 1-2) to late phase (trials 9-10) for each axis- x (left-right), y (up-down), and z (front-back) (related to Figure 2C).

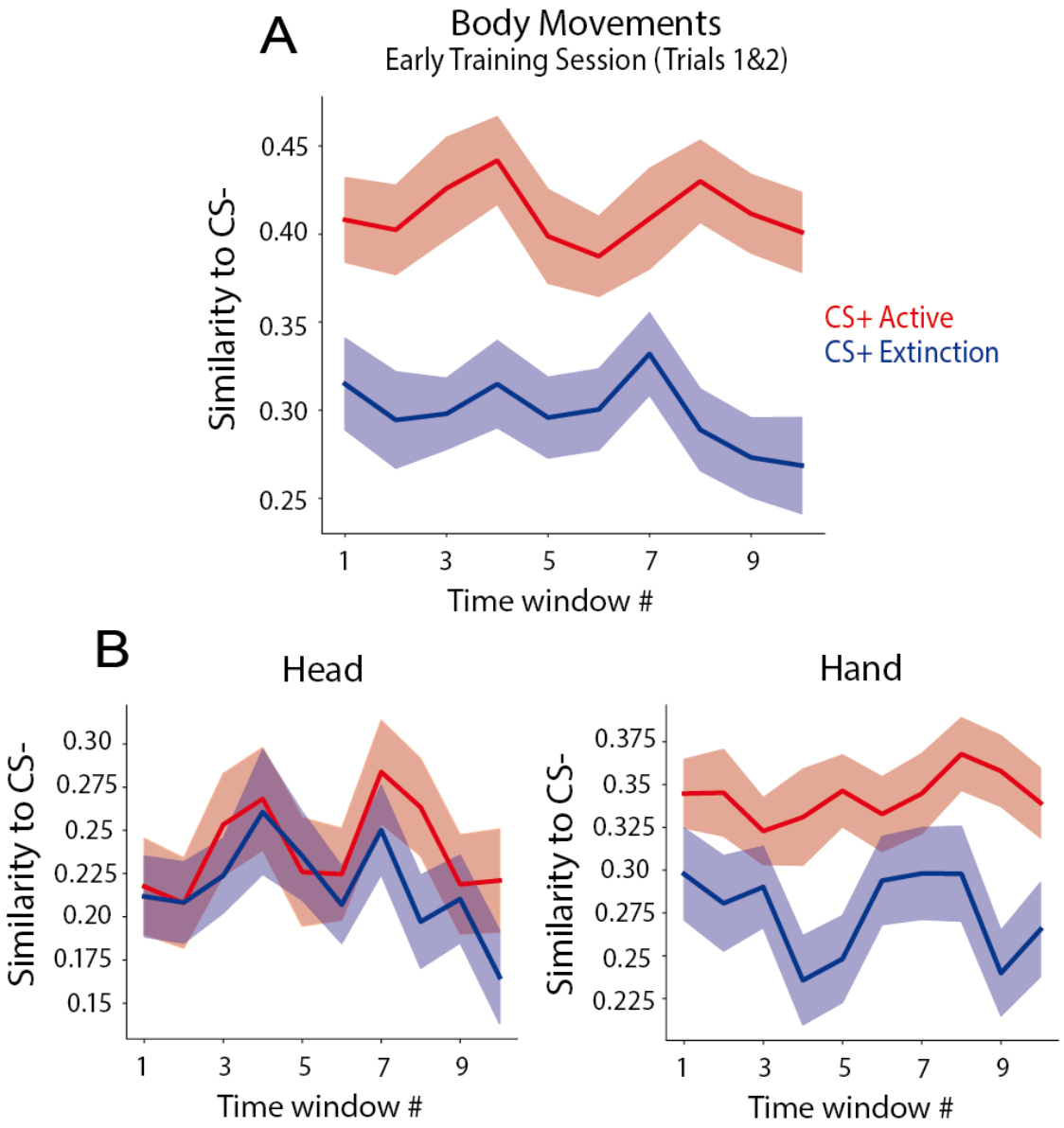

Figure S3. Body movement patterns during the ACT training session. (A) At the start of the training (trials 1-2), similarity in body movement patterns toward the CS+ versus the CS- avatars already deviated between the ACT and EXT condition in all time windows (all two-tailed t-values(40)> 6.43, p-values<0.001); (B) Exploring the movement pattern similarity for hand and head separately corroborated that different task requirements and task instructions (i.e., using spray in the ACT condition) confounded the movement pattern analysis during training (related to Figure 3C).

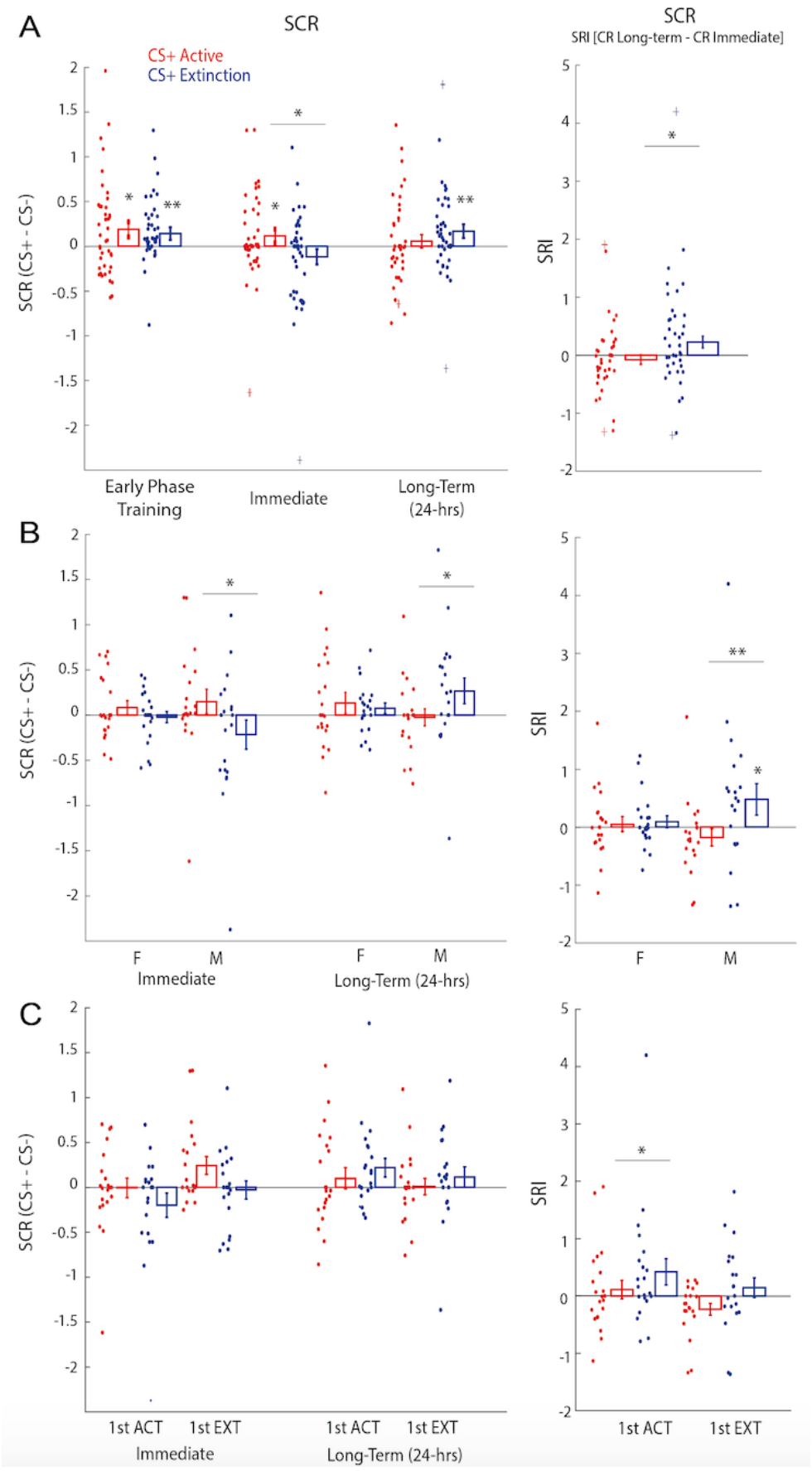
A
SCR
CS+ Active
CS+ Extinction
SCR (CS+ - CS-)
Early Phase Training
Immediate
Long-Term (24-hrs)
SCR
SRI [CR Long-term - CR Immediate]
SRI
B
F
M
Immediate
Long-Term (24-hrs)
C
1st ACT
1st EXT

Figure S4. (A) Conditioned responses in SCR after removing outliers (>=3SD) in Experiment 1. Conditioned responses in SCR during the early phase of the training sessions and Immediate and long-term effects of ACT and EXT training (left), Spontaneous Recovery Index (SRI), as measured by the difference between the immediate versus the long-term conditioned SCR (right). All outliers were defined based on the CR values [CS+ - CS-]. (B,C) Effects of gender (M: Male, F: Female) and training order on ACT and EXT training Immediate and long-term effects (left) and on SRI (right). Main results supporting the emergence of CR 24hrs after the conditioning session [one-tailed CS+ > CS- ; $CS+_{active}$ $t(39) = 1.97$, $p = 0.02$; BF+0 = 1.84; $CS+_{extinction}$ $t(39) = 3.22$, $p = 0.001$; BF+0 = 26.42; $CS+_{active}$ Vs $CS+_{extinction}$ $t(39) = 0.39$, $p = 0.69$; BF01 = 5.44] and the more robust long-term reduction of CR to $CS+_{active}$ relative to $CS+_{extinction}$ [SRI to $CS+_{extinction}$ Vs SRI to $CS+_{active}$; $t(38) = -2.32$, $p = 0.01$; BF+0 = 3.70] remained quantitatively similar after removing outliers (3 SD apart from mean) (Fig S4A) (related to Figure 3B).

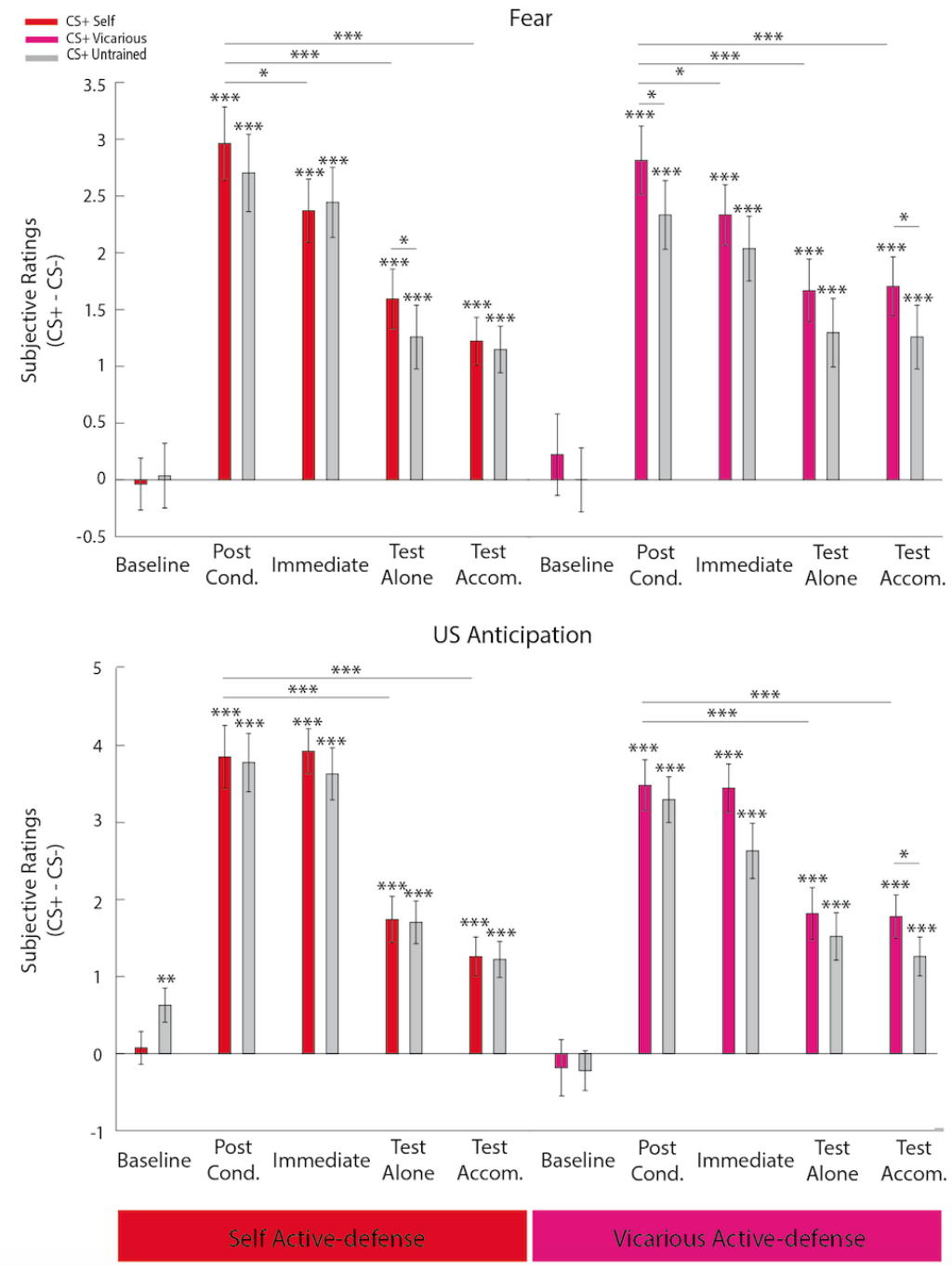

Figure S5. Comparison of fear alleviation effects between Self and Vicarious ACT training on subjective ratings. Conditioned responses [CS+ - CS-] in subjective ratings of fear (left) and US-anticipation (right) before and after conditioning on day 1, after ACT training on day 1 and after both Test-Alone and Test-Accompanied on day 2 [one-tailed CS+ > CS-; Self ACT; Fear: $CS+_{active}$ $t(26) = 9.19$, $p < 0.001$; BF+0 >100; $CS+_{untrained}$ $t(26) = 7.29$, $p < 0.001$; BF+0 >100, US likelihood: $CS+_{active}$ $t(26) = 9.41$, $p < 0.001$; BF+0 >100; $CS+_{untrained}$ $t(26) = 9.97$, $p < 0.001$; BF+0 >100; Vicarious ACT; Fear: $CS+_{active}$ $t(26) = 9.31$, $p < 0.001$; BF+0 >100; $CS+_{untrained}$ $t(26) = 7.72$, $p < 0.001$; BF+0 >100, US likelihood: $CS+_{active}$ $t(26) = 10.53$, $p < 0.001$; BF+0 >100; $CS+_{untrained}$ $t(26) = 11.12$, $p < 0.001$; BF+0 >100] (related to Figure 2 and 4).

Acknowledgment
The project was supported by JST Presto (18068712), JST Moonshot (20343198) and JSPS B grants (18H02714 and 22H01111) awarded to AK. We thank Yuna Ujiie and Kazuto Ishii for their support in conducting experiments, and Hikaru Yokoyama and Naotsugu Kaneko for their support with the equipments. We appreciate Anikó Kusztor for her helpful comments on the manuscript.

References

Adolphs, R. (2013). The biology of fear. *Current Biology : CB*, *23*(2). https://doi.org/10.1016/J.CUB.2012.11.055

Amat, J., Baratta, M. V., Paul, E., Bland, S. T., Watkins, L. R., & Maier, S. F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience 2005 8:3*, *8*(3), 365–371. https://doi.org/10.1038/nn1399

An, B., Kim, J., Park, K., Lee, S., Song, S., & Choi, S. (2017). Amount of fear extinction changes its underlying mechanisms. *ELife*, *6*. https://doi.org/10.7554/ELIFE.25224

An, X., Zhang, F., Liu, Y., Yang, P., & Yu, D. (2019). Remote fear memory is sensitive to reconditioning. *Behavioural Brain Research*, *359*, 723–730. https://doi.org/10.1016/J.BBR.2018.09.013

Arnold, M. B. (1960). *Emotion and personality. - PsycNET*. Retrieved from https://psycnet.apa.org/record/1960-35012-000

Baas, J. M., Nugent, M., Lissek, S., Pine, D. S., & Grillon, C. (2004). Fear conditioning in virtual reality contexts: A new tool for the study of anxiety. *Biological Psychiatry*, *55*(11), 1056–1060. https://doi.org/10.1016/J.BIOPSYCH.2004.02.024

Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology*, *92*(1), 90–96. https://doi.org/10.1016/J.BIOPSYCHO.2011.12.015

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Bishop, S., Duncan, J., Brett, M., & Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. *Nature Neuroscience*, *7*(2), 184–188. https://doi.org/10.1038/NN1173

Boeke, E. A., Moscarello, J. M., LeDoux, J. E., Phelps, E. A., & Hartley, C. A. (2017). Active Avoidance: Neural Mechanisms and Attenuation of Pavlovian Conditioned Responding. *Journal of Neuroscience*, *37*(18), 4808–4818. https://doi.org/10.1523/JNEUROSCI.3261-16.2017

Bouton, M. E., & Bolles, R. C. (1980). Conditioned fear assessed by freezing and by the suppression of three different baselines. *Animal Learning & Behavior 1980 8:3*, *8*(3), 429–434. https://doi.org/10.3758/BF03199629

Brown, J. S., Kalish, H. I., & Farber, I. E. (1951). Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *Journal of Experimental Psychology*, *41*(5), 317–328. https://doi.org/10.1037/H0060166

Cain, C. K., & LeDoux, J. E. (2007). Escape from fear: a detailed behavioral analysis of two atypical responses reinforced by CS termination. *Journal of Experimental Psychology. Animal Behavior Processes*, *33*(4), 451–463. https://doi.org/10.1037/0097-7403.33.4.451

Cisler, J. M., Begle, A. M., Amstadter, A. B., Resnick, H. S., Danielson, C. K., Saunders, B. E., & Kilpatrick, D. G. (2012). Exposure to interpersonal violence and risk for PTSD, depression, delinquency, and binge drinking among adolescents: data from the NSA-R. *Journal of Traumatic Stress*, *25*(1), 33–40. https://doi.org/10.1002/JTS.21672

Cohen, D. H., & Randall, D. C. (2003). Classical Conditioning of Cardiovascular Responses. *Http://Dx.Doi.Org/10.1146/Annurev.Ph.46.030184.001155*, *VOL. 46*, 187–197. https://doi.org/10.1146/ANNUREV.PH.46.030184.001155

Collins, K. A., Mendelsohn, A., Cain, C. K., & Schiller, D. (2014). Taking Action in the Face of Threat: Neural Synchronization Predicts Adaptive Coping. *The Journal of Neuroscience*, *34*(44), 14733. https://doi.org/10.1523/JNEUROSCI.2152-14.2014

Davis, M. (1990). Animal models of anxiety based on classical conditioning: the conditioned

emotional response (CER) and the fear-potentiated startle effect. *Pharmacology & Therapeutics*, *47*(2), 147–165. https://doi.org/10.1016/0163-7258(90)90084-F

De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews. Neuroscience*, *7*(3), 242–249. https://doi.org/10.1038/NRN1872

Delgado, M. R., Jou, R. L., LeDoux, J. E., & Phelps, E. A. (2009). Avoiding negative outcomes: Tracking the mechanisms of avoidance learning in humans during fear conditioning. *Frontiers in Behavioral Neuroscience*, *3*(NOV), 33. https://doi.org/10.3389/NEURO.08.033.2009/BIBTEX

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 of the United States of America*, *108*(28), 11721–11726. https://doi.org/10.1073/PNAS.1108239108/SUPPL_FILE/PNAS.201108239SI.PDF

Eldar, E., Hauser, T. U., Dayan, P., & Dolan, R. J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4812–4817. https://doi.org/10.1073/PNAS.1519829113/SUPPL_FILE/PNAS.1519829113.SAPP.PDF

Evans, D. A., Stempel, A. V., Vale, R., & Branco, T. (2019). Cognitive Control of Escape Behaviour. *Trends in Cognitive Sciences*, *23*(4), 334–348. https://doi.org/10.1016/J.TICS.2019.01.012/ATTACHMENT/976B07EC-38BF-48EB-BCEE-1DDC436E8B4A/MMC1.PDF

Fanselow, M. (1988). A functional behavioristic approach to aversively motivated behavior: Predatory imminence as a determinant of the topography of defensive behavior. Retrieved June 17, 2022, from https://www.semanticscholar.org/paper/A-functional-behavioristic-approach-to-aversively-a-Fanselow-Lester/2d798ca231935e0948dc67936619864725211cff

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 2007 39:2*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Faul, L., Stjepanovic, D., Stivers, J. M., Stewart, G. W., Graner, J. L., Morey, R. A., & LaBar, K. S. (2020). Proximal threats promote enhanced acquisition and persistence of reactive fear-learning circuits. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(28), 16678–16689. https://doi.org/10.1073/PNAS.2004258117/SUPPL_FILE/PNAS.2004258117.SAPP.PDF

Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2007). Poly-victimization: a neglected component in child victimization. *Child Abuse & Neglect*, *31*(1), 7–26. https://doi.org/10.1016/J.CHIABU.2006.06.008

Frijda, N. H. (1986). *The emotions*. Retrieved from https://psycnet.apa.org/record/1987-97938-000

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., Selbing, I., Flygare, O., Öhman, 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

Grillon, C., Baas, J. M. P., Cornwell, B., & Johnson, L. (2006). Context conditioning and behavioral avoidance in a virtual reality environment: effect of predictability. *Biological Psychiatry*, *60*(7), 752–759. https://doi.org/10.1016/J.BIOPSYCH.2006.03.072

Gruene, T. M., Flick, K., Stefano, A., Shea, S. D., & Shansky, R. M. (2015). Sexually divergent expression of active and passive conditioned fear responses in rats. *ELife*, *4*(NOVEMBER2015). https://doi.org/10.7554/ELIFE.11352.001

Haaker, J., Diaz-Mataix, L., Guillazo-Blanch, G., Stark, S. A., Kern, L., LeDoux, J. E., & Olsson, A. (2021). Observation of others' threat reactions recovers memories previously shaped by firsthand experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(30). https://doi.org/10.1073/PNAS.2101290118/SUPPL_FILE/PNAS.2101290118.SAPP.PDF

Hartley, C. A., Coelho, C. A. O., Boeke, E., Ramirez, F., & Phelps, E. A. (2019). Individual differences in blink rate modulate the effect of instrumental control on subsequent Pavlovian responding. *Psychopharmacology*, *236*(1), 87–97. https://doi.org/10.1007/S00213-018-5082-6/FIGURES/4

Hartley, C. A., Gorun, A., Reddan, M. C., Ramirez, F., & Phelps, E. A. (2014). Stressor controllability modulates fear extinction in humans. *Neurobiology of Learning and Memory*, *113*, 149–156. https://doi.org/10.1016/J.NLM.2013.12.003

Keum, S., & Shin, H. S. (2019). Neural Basis of Observational Fear Learning: A Potential Model of Affective Empathy. *Neuron*, *104*(1), 78–86. https://doi.org/10.1016/J.NEURON.2019.09.013

Kilpatrick, D. G., Ruggiero, K. J., Acierno, R., Saunders, B. E., Resnick, H. S., & Best, C. L. (2003). Violence and Risk of PTSD, Major Depression, Substance Abuse/Dependence, and Comorbidity: Results From the National Survey of Adolescents, *71*(4), 692–700. https://doi.org/10.1037/0022-006X.71.4.692

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., … Lau, H. (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature Human Behaviour 2016 1:1*, *1*(1), 1–7. https://doi.org/10.1038/s41562-016-0006

Kozlowska, K., Walker, P., McLean, L., & Carrive, P. (2015). Fear and the Defense Cascade: Clinical Implications and Management. *Harvard Review of Psychiatry*, *23*(4), 263. https://doi.org/10.1097/HRP.0000000000000065

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 4. https://doi.org/10.3389/NEURO.06.004.2008/BIBTEX

Lang, P. J. (1995). The Emotion Probe: Studies of Motivation and Attention. *American Psychologist*, *50*(5), 372–385. https://doi.org/10.1037/0003-066X.50.5.372

Ledoux, J., & Daw, N. D. (2018). Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews. Neuroscience*, *19*(5), 269–282. https://doi.org/10.1038/NRN.2018.22

LeDoux, J. E. (2019). *The deep history of ourselves : the four-billion-year story of how we got conscious brains*.

LeDoux, J. E., Moscarello, J., Sears, R., & Campese, V. (2017). The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm. *Molecular Psychiatry*, *22*(1), 24–36. https://doi.org/10.1038/MP.2016.166

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., … Merz, C. J. (2017). Don't fear "fear conditioning": Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience and Biobehavioral Reviews*, *77*, 247–285. https://doi.org/10.1016/J.NEUBIOREV.2017.02.026

Lu, Z., & Ku, Y. (2020). NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. *Frontiers in Neuroinformatics*, *14*. https://doi.org/10.3389/FNINF.2020.563669

Ly, A., Verhagen, J., & Wagenmakers, E. J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/J.JMP.2015.06.004

Martinez, R. C. R., Gupta, N., Lazaro-Munoz, G., Sears, R. M., Kim, S., Moscarello, J. M., …

Cain, C. K. (2013). Active vs. reactive threat responding is associated with differential c-Fos expression in specific regions of amygdala and prefrontal cortex. *Learning & Memory*, *20*(8), 446–452. https://doi.org/10.1101/LM.031047.113

Meulders, A., & Vlaeyen, J. W. S. (2012). Reduction of fear of movement-related pain and pain-related anxiety: An associative learning approach using a voluntary movement paradigm. *PAIN*, *153*(7), 1504–1513. https://doi.org/10.1016/J.PAIN.2012.04.013

Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., … Frith, C. D. (2009). From Threat to Fear: The Neural Organization of Defensive Fear Systems in Humans. *The Journal of Neuroscience*, *29*(39), 12236. https://doi.org/10.1523/JNEUROSCI.2378-09.2009

Myers, K. M., & Davis, M. (2007). Mechanisms of fear extinction. *Molecular Psychiatry*, *12*(2), 120–150. https://doi.org/10.1038/SJ.MP.4001939

Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *13*(6), 681–685. https://doi.org/10.1101/LM.393906

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., & Chon, K. H. (2018). Electrodermal Activity Is Sensitive to Cognitive Stress under Water. *Frontiers in Physiology*, *8*(JAN). https://doi.org/10.3389/FPHYS.2017.01128

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, *22*(3), 561–569. https://doi.org/10.1016/J.JANXDIS.2007.04.006

Reichenberger, J., Pfaller, M., Forster, D., Gerczuk, J., Shiban, Y., & Mühlberger, A. (2019). Men scare me more: Gender differences in social fear conditioning in virtual reality. *Frontiers in Psychology*, *10*(JULY), 1617. https://doi.org/10.3389/FPSYG.2019.01617/BIBTEX

Robbins, S. J. (1990). Mechanisms Underlying Spontaneous Recovery in Autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*(3), 235–249. https://doi.org/10.1037/0097-7403.16.3.235

Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718). https://doi.org/10.1098/RSTB.2016.0206

Rösler, L., & Gamer, M. (2019). Freezing of gaze during action preparation under threat imminence. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-53683-4

Slater, M., Neyret, S., Johnston, T., Iruretagoyena, G., Crespo, M. Á. de la C., Alabèrnia-Segura, M., … Feixas, G. (2019). An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. *Scientific Reports 2019 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-46877-3

Steckle, L. C. (2010). A Trace Conditioning of the Galvanic Reflex. *Http://Dx.Doi.Org/10.1080/00221309.1933.9920953*, *9*(2), 475–480. https://doi.org/10.1080/00221309.1933.9920953

Tada, M., Uchida, H., Maeda, T., Konishi, M., Umeda, S., Terasawa, Y., … Takahashi, T. (2015). Fear Conditioning Induced by Interpersonal Conflicts in Healthy Individuals. *PLOS ONE*, *10*(5), e0125729. https://doi.org/10.1371/JOURNAL.PONE.0125729

Taschereau-Dumouchel, V., Kawato, M., & Lau, H. (2019). Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Molecular Psychiatry 2019 25:10*, *25*(10), 2342–2354. https://doi.org/10.1038/s41380-019-0520-3

Thompson, A., McEvoy, P. M., & Lipp, O. V. (2018). Enhancing extinction learning: Occasional presentations of the unconditioned stimulus during extinction eliminate spontaneous recovery, but not necessarily reacquisition of fear. *Behaviour Research and Therapy*, *108*, 29–39. https://doi.org/10.1016/J.BRAT.2018.07.001

Trigili, E., Grazi, L., Crea, S., Accogli, A., Carpaneto, J., Micera, S., … Panarese, A. (2019). Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks. *Journal of NeuroEngineering and Rehabilitation*, *16*(1). https://doi.org/10.1186/S12984-019-0512-1

Vaitl, D., & Lipp, O. V. (1997). Latent inhibition and autonomic respones: A psychophysiological approach. *Behavioural Brain Research*, *88*(1), 85–93. https://doi.org/10.1016/S0166-4328(97)02310-3

Vervliet, B., Baeyens, F., Van den Bergh, O., & Hermans, D. (2013). Extinction, generalization, and return of fear: a critical review of renewal research in humans. *Biological Psychology*, *92*(1), 51–58. https://doi.org/10.1016/J.BIOPSYCHO.2012.01.006

Visser, R. M., Kunze, A. E., Westhoff, B., Scholte, H. S., & Kindt, M. (2015). Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. *Psychoneuroendocrinology*, *55*, 8–20. https://doi.org/10.1016/J.PSYNEUEN.2015.01.021

Wang, W., Schuette, P. J., Nagai, J., Tobias, B. C., Fernando, F. M., Ji, S., … Adhikari, A. (2021). Coordination of escape and spatial navigation circuits orchestrates versatile flight from threats. *Neuron*, *109*(11), 1848-1860.e8. https://doi.org/10.1016/J.NEURON.2021.03.033

Wanke, N., & Schwabe, L. (2020). Dissociable neural signatures of passive extinction and instrumental control over threatening events. *Social Cognitive and Affective Neuroscience*, *15*(6), 625–634. https://doi.org/10.1093/SCAN/NSAA074

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1–14. https://doi.org/10.1037/H0069608