HHS Public Access

Author manuscript

Neuropsychologia. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:

Neuropsychologia. 2020 November ; 148: 107653. doi:10.1016/j.neuropsychologia.2020.107653.

Generalization of conditioned fear along a dimension of increasing positive valence

Mason McClay[1], **Augustin C. Hennings**[2], **Alex Reidel**[1], **Joseph E. Dunsmoor**[1,2]

[1]Department of Psychiatry, Dell Medical School, University of Texas at Austin

[2]Institute for Neuroscience, University of Texas at Austin

Abstract

The amount of fear evoked by potential threats is oftentimes proportional to the overlap in shared features with known threats. An adaptive learning system should therefore extract relevant features from threat stimuli to successfully detect other novel threats in the environment. But what if the most relevant feature of a threat stimulus is emotionally positive? Here, we used Pavlovian fear conditioning to ask whether people extract positive emotional features of a fear conditioned stimulus (CS) to selectively generalize to other stimuli that contain positive features. In a between subjects design, we first paired a picture of a face expressing either a slight amount of happiness or fear with an electrical shock to the wrist. We then tested fear generalization to modified face stimuli of the same identity expressing more or less happiness or fear. Both groups exhibited biased physiological arousal (a peak shift) to a face stimulus with the most exaggerated emotional expression, regardless of valence. Fear generalization diminished to unreinforced happy faces over the course of testing, whereas arousal was maintained to unreinforced fearful faces throughout testing. Finally, subjects fear conditioned to a slightly happy face were accurate at retrospectively identifying the correct CS, whereas subjects fear conditioned to a fearful face retrospectively misidentified a more fearful face as the threat CS. These findings suggest that overlap of positive emotional features extracted from a known threat can guide biased fear generalization, but that generalization is maintained by an intensity-based dimension of increasing fear and produces retrospective biases in threat intensity estimation.

Keywords

emotion learning; fear; fear generalization; generalization learning; Pavlovian conditioning

Correspondence: Joseph E Dunsmoor, joseph.dunsmoor@austin.utexas.edu.

Author Contributions.

MM and JED developed and designed the experiments. MM, AR, and AH conducted the data collection and analyzed the results. MM, AH, and JED wrote the manuscript.

Disclosure Statement.

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Data Availability Statement.

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

INTRODUCTION

When confronted with new threats, animals can adapt behaviors that aid in survival by relying on prior experiences with similar threats (Dunsmoor, Prince, Murty, Kragel, & LaBar, 2011; Onat & Büchel, 2015). Accordingly, an animal should extract features from a known threat that most reliably predicts danger, so as to respond appropriately to perceptually similar stimuli in the future. Behavioral generalization based on perceptual features can occur along routes of similarity (e.g., similar shapes or colors) and intensity (e.g., brightness or volume) (Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015). Similarity-based generalization is strongest to stimuli that most closely resemble the conditioned stimulus (CS), and diminishes as similarity decreases (Guttman & Kalish, 1956; Vervliet, Kindt, Vansteenwegen, & Hermans, 2010). In contrast, intensity-based generalization is biased toward stimuli that are more intense than the CS, producing an asymmetrical shift in maximal responses from the CS (e.g., a dim light) toward stimuli that possess more intense features (e.g., a brighter light) (Ghirlanda & Enquist, 2003). Real-world stimuli, however, are complex and contain features that can be represented along dimensions of similarity, intensity, and even valence.

Importantly, the degree to which a potential threat is regarded as dangerous is likely based on which feature of a complex stimulus is most associated with danger. Prior research shows that fear generalization is more sensitive to the amount of negative emotional intensity in a face (a face expressing slight fear) than the overlap in perceptual similarity (actor identity), leading to biased generalization toward a face of the same identity expressing more fear, as compared to the same face expressing less fear (Dunsmoor, Mitroff, & LaBar, 2009). Here, we examined whether fear generalization is sensitive to the amount of *positive* emotional intensity in a face (happiness). Specifically, we test a counterintuitive proposal that fear acquired to a face expressing a slight smile produces an asymmetrical shift in maximal fear responses to faces expressing even more happiness, as compared to the same face identity expressing less happiness.

Consider an example in which an individual has a terrifying encounter with a dog and then develops a widespread fear of all dogs. Despite a generalized fear of dogs, some dogs are likely to elicit more fear (e.g., a large German Shepard) than others (e.g., a little Pug) (Dymond et al., 2015). In this scenario, the original threatening dog contained some features considered phylogenetically threatening (e.g. sharp teeth, darting movements, aggressive vocalizations) as well as typical dog-like features (e.g. floppy ears). When assessing the threat value of other dogs, fear-relevant features associated with danger (e.g., size of its teeth) are likely prioritized over those less fear-relevant features (e.g., floppiness of its ears), and thus determine generalization to other category members who share some combination of features associated with the original threat. According to fear preparedness theory (Ohman & Mineka, 2001), the neural circuitry responsible for defensive behavior and fear phenomenology evolved to be selective toward these types of ecologically threatening features. Indeed, several studies have demonstrated that evolutionarily fear-relevant stimuli (e.g., snakes or spiders) are more readily associated with an aversive outcome and take

longer to extinguish than threat-irrelevant stimuli (Ohman & Mineka, 2001; Seligman, 1970; but see Åhs et al., 2018 for a systematic review critiquing prepareness theory).

However, recent work suggests fear conditioning is selective to features that are merely *relevant* to an animals' concerns, irrespective of valence, and is therefore not driven purely by fear-relevant features. This includes goal-relevant (Stussi, Ferrero, Pourtois, & Sander, 2019) and even positive-relevant stimuli (e.g., erotic images) (Stussi, Pourtois, & Sander, 2018a), in line with the idea that emotional learning and behavior is honed by a more general relevance detection system (Sander, Grafman, & Zalla, 2003). Differing accounts for selective fear conditioning to fear-relevant or merely *relevant* CSs leads to different predictions for whether fear generalization is sensitive to the positively valenced features of a learned threat (e.g., generalizing fear from a cute dog that attacked you to an even cuter dog). If fear conditioning is only sensitive to negative fear-relevant features, then learning to fear a CS with a positive emotional feature should produce similarity-based generalization that peaks at the CS and diminishes as physical similarity to the CS decreases. If, however, fear conditioning is selective to the most relevant feature, regardless of valence, then generalization should resemble an intensity gradient with maximal fear responses to unreinforced stimuli that share a more extreme positive feature associated with the original CS.

Accordingly, we compared behavioral generalization gradients to face stimuli bearing positive or negative features that can be represented along a dimension of increasing intensity, specifically emotional expression. We used the same actor face morphed between neutral and emotional endpoints in order to simultaneously assess similarity-based and intensity-based generalization using the same stimulus dimension (Dunsmoor & LaBar, 2013; Dunsmoor et al., 2009). Given the importance of avoiding inherently threatening stimuli in the natural world, it is perhaps counterintuitive that *fear* generalization would be sensitive to the amount of *happiness* in a novel (unreinforced) stimulus. Thus, one possibility is that maximal fear responses (assessed here with physiological arousal) should be maximal to the known threat, and diminish as similarity decreases, such that the happiest face will elicit little arousal. In contrast, a domain general mechanism of relevance detection might support skewed behavioral generalization to unreinforced stimuli that contain any salient feature with the CS that can be represented along a dimension of increasing intensity, irrespective of the valence of the feature. A domain general mechanism would produce intensity-based generalization gradients along dimensions of both increasing happiness and fear.

Participants in two groups were fear conditioned to a face morphed along a dimension of either happiness (Happy Group) or fear expression (Fear Group). Fear conditioning was then followed by a fear generalization test to a gradient of faces morphed along the respective facial expression containing more or less emotional expression intensity than the original threat CS. Given evidence supporting the capacity for enhanced fear learning to positive emotional stimuli (Stussi et al., 2018a), as well as intensity-based accounts of stimulus generalization (Ghirlanda & Enquist, 2003), we predicted conditioned fear responses would generalize as a function of emotional intensity, regardless of valence. But importantly new learning occurs over the course of generalization testing as subjects realize that test stimuli

are not reinforced. As such, we also predicted that fear generalization along a dimension of positive emotional expression would diminish over the course of testing such that generalization to happy faces would diminish more rapidly than generalization to fearful faces. Finally, prior work assessing retrospective memory for the specific CS identity after fear generalization testing shows biased selection of a more exaggerated fearful face (Dunsmoor et al., 2009, 2011; Morey et al., 2015). Therefore, we predicted that participants fear conditioned to a slightly fearful face would retrospectively misremember the CS as more intense than it actually was, but that this memory bias would be less pronounced when the original threat was a happy face.

METHODS

Participants.

Forty healthy adults (29 female: M = 25.30 years, SD = 3.23) participated in one of the two groups described below. One participant failed to finish Group 1 (n = 19) and equipment failure occurred while running one participant in Group 2 (n = 19). The target sample size of 40 was determined by a power analysis performed in G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). We used a large effect size from a main effect of Stimulus Intensity (within-group term) from prior research (η_p^2 = .27; Dunsmoor et al., 2009) to estimate a moderate effect size for a between-within interaction term (Group by Stimulus Intensity; η_p^2 = .07) with target parameters of power = .80 and alpha = .05. Our sample size yielded enough power for a between-within interaction (Group by Stimulus Intensity). Recruitment exclusionary criteria included a self-reported history of psychiatric, neurological, or major medical illnesses or current use of psychoactive medication. The study was approved by the Institutional Review Board at the University of Texas at Austin (IRB #2017-02-0094).

Stimulus Set.

A male face morphed along a gradient from neutral-to-happy was used as the stimulus set for Happy Group and the same male face morphed along a gradient from neutral-to-fearful was used for Fear Group. This identity was taken from the Ekman pictures of facial affect (Ekman & Friesen, 1976) and was morphed along a continuum from neutral-to-happy and neutral-to-fearful (LaBar, Crupain, Voyvodic, & McCarthy, 2003). Face morphs were positioned in a full-frontal orientation and cropped to include only the face, without hair, ears, or neckline. Images were normalized for contrast and luminance and appeared on a gray rectangular background.

Participants were exposed to the same face identity during all experimental phases to ensure that only happy and fear expression and no other features related to actor identity (i.e., markings specific to an individual face such as moles or facial hair), were manipulated. Six faces along the happy continuum were used for the Happy Group: 100% neutral, 11% happy/88% neutral, 33% happy/66% neutral, 55% happy/44% neutral, 77% happy/22% neutral, and 100% happy; and six faces along the fear continuum were used for the Fear Group: 100% neutral, 11% fear/88% neutral, 33% fear/66% neutral, 55% fear/44% neutral, 77% fear/22% neutral, and 100% fear. For the Happy Group, these stimuli are also labeled as SN, SH1, SH2, SH3, SH4, and SH5, respectively, where SN refers to ‘Stimulus Neutral’

and SH refers to ‘Stimulus Happy.’ Likewise, for the Fear Group, these stimuli were labelled as SN, SF1, SF2, SF3, SF4, and SF5, where SF refers to ‘Stimulus Fear.’ Hence, either the SH3 (Happy Group) or SF3 (Fear Group) served as the conditioned stimulus (CS+), and SN served as the unpaired CS (CS−) during fear conditioning. Importantly, both groups used the same neutral face (SN) as the unpaired CS-during fear conditioning.

Skin conductance.

Skin conductance was recorded using a BIOPAC MP160 system (BIOPAC Systems Inc.; Santa Barbara, CA) and sampled at 200 Hz. Pre-gelled BIOPAC EL507 disposable snap electrodes were attached to the hypothenar eminence of the palmar surface of the left hand. Skin conductance responses (SCR) for each phase of the study were calculated using a Matlab script (Autonomate; Green, Kragel, Fecteau, & LaBar, 2014) and hand-scored by the third author for validity and to check for potential artefacts. SCRs were considered related to stimulus presentation if the trough-to-peak response (1) occurred 1–4 s following stimulus onset, (2) lasted between 0.5 and 5.0 s, and (3) was greater than 0.02 microsiemens (μS). If these criteria were not met, the SCR was scored as zero. SCRs were square-root transformed to normalize the distribution.

Electrical shock procedure.

An electrical shock (50 milliseconds) delivered to the right wrist served as the unconditioned stimulus for both groups. Electrical shocks were delivered via electrodes connected to a constant voltage stimulator (STM 200, BIOPAC Systems). The intensity of the shock was calibrated for each participant in a stepwise manner, starting from a low barely perceptible setting and ending at a level that the participant rated as being “highly annoying and unpleasant, but not painful”, based on the protocols of prior threat conditioning studies (Dunsmoor et al., 2014).

Task Design and Procedure.

After obtaining informed consent and attaching the electrodes, participants viewed a 5-minute video that served as a wash-out prior to the start of the conditioning experiment. Videos were obtained from an episode of Norsk Slow TV (Neiderer, 2017). Participants were instructed to turn off their phones. Participants were told that the video was unrelated to the experiment and to simply pay attention for five minutes.

During the fear conditioning phase, participants were presented with both the CS+ (SH3; 55% happy in the Happy Group, SF3; 55% fearful in the Fear Group) and the CS− (SN; 100% neutral in both groups) 15 times each. Each face was presented for 4.5 seconds and was separated by an ITI of a fixation cross on a blank screen that lasted 6 seconds. Faces were presented in a pseudorandom order and co-terminated with the US in 9 of the 15 CS+ presentations (60% CS+US reinforcement rate). Following fear conditioning, participants had a five-minute break during which another segment of the neutral video was shown, simply so that each participant engaged in the same task-irrelevant activity during the short break. Participants then underwent the fear generalization test phase, during which they were presented with each of the 5 faces nine times (45 trials total). Each trial was separated by a 6 second ITI. The CS+(SH3 or SF3) was intermittently paired with the US on 33% of

generalization test trials (“steady-state” generalization test; Honig & Urcuioli, 1981). These steady-state (“booster” trials) reinforcement procedures are intended to extend the length of time over which responding can be measured and offset the effects of extinction and habituation (Honig & Urcuioli, 1981; Lim & Pessoa, 2008; Mednick & Freedman, 1960). Generalization was tested either to faces along a continuum of happiness (Happy Group) or fear (Fear Group). As a slight modification to our prior design (Dunsmoor et al., 2009), the CS− was not included during the generalization test. This allows us to symmetrically test generalization for morphed faces along a gradient of increasing emotional intensity, with two stimuli of lesser and two of greater emotional intensity than the CS+. On each trial of fear conditioning and the generalization test, participants were instructed to indicate if the face was displaying either happiness (Happy Group) or fear (Fear Group) or not by pressing 1 (no) or 2 (yes) on the keyboard.

Statistical analysis.

Statistical analyses were performed in python utilizing the pingouin package (Vallat, 2018), the SciPy package (Jones Oliphant, T., Peterson, P., SciPy community, 2001), as well as R for mixed-effects regression modelling (lmer function in the lme4 library; Bates D, Mächler M, Bolker B, & Walker S, 2015). Behavioral data from fear conditioning, SCRs and reaction times (RTs), were analyzed using a two-way analysis of variance (ANOVA), with within-subjects factors of Stimulus Intensity (CS+/−) and a between-subjects factor of Group (Happy/Fear). Within each group, bootstrapped 95% confidence intervals (CI; *k iterations* = 5000) were computed to determine differences of SCR and reaction times between CS− and CS+ during fear conditioning, with CIs that do not contain zero within their range considered evidence of a significant effect.

To investigate the temporal dynamics across the generalization test, behavioral data were binned into 3 equal tertiles of 15 trials each (see Dunsmoor et al., 2009). Face stimuli were counterbalanced within each tertile. Fear generalization of SCRs and RTs was analyzed using a three-way ANOVA with within-subjects factors of stimulus intensity and tertile, and a between-subjects factor of group. The temporal dynamics of fear generalization were further investigated using mixed-effects models. The goal of these tests was to determine whether a model which contained a quadratic fit, as opposed to linear, could better explain the peak-shift of SCRs observed during generalization test. Mixed-effect models were fit to SCRs for face stimuli during the generalization test and included subject as a random factor. Likelihood ratio tests were used to determine if the quadratic models significantly outperformed linear models (Vanbrabant et al., 2015). A Kruskal-Wallis test was used to investigate differences in retrospective CS+ identification across groups.

RESULTS

Fear conditioning

SCR analysis—A two-way ANOVA yielded a significant main effects of Stimulus Intensity ($F(1, 36) = 87.803$, $p < .01$, $\eta_p^2 = .709$) and Group ($F(1, 36) = 14.143$, $p < .01$, $\eta_p^2 = .282$), as well as a significant Stimulus Intensity by Group interaction ($F(1, 36) = 21.676$, $p < .01$, $\eta_p^2 = .376$). Successful fear conditioning was verified by heightened SCRs

to the CS+ versus the CS− in both groups (Happy Group: paired $t(18) = 3.3$, $p < .01$, 95% CI = [.074, .245], Figure 1a; Fear Group: paired $t(18) = 10$, $p < .01$, 95% CI = [0.36, 0.53], Figure 1b). Thus, while differential fear conditioning was greater in the Fear Group, both groups exhibited successful discriminative fear conditioning.

Reaction times and emotion ratings of CS+ and CS−—A two-way ANOVA yielded a significant main effect of Stimulus Intensity ($F(1, 36) = 9.171$, $p < .01$, $\eta_p^2 = .203$). There was no significant main effect of Group ($F(1, 36) = .348$, $p = .56$, $\eta_p^2 = .01$), or Stimulus Intensity by Group interaction ($F(1, 36) = .263$, $p < .61$, $\eta_p^2 = .007$). Planned comparisons showed that the Happy Group exhibited faster RTs on CS+ versus CS− trials (paired $t(18) = 2.63$, $p = .017$, 95% CI = [−27.97, −169.68]; Figure 2a). The Fear Group exhibited nominally faster RTs on CS+ versus CS− trials, but this difference was not significant (paired $t(18) = 1.705$, $p = .105$, 95% CI [−152.87, 0.29]; Figure 2b). This pattern of RT results was somewhat surprising given our prior findings that RTs were faster for the neutral versus quasi-fearful face during fear conditioning (Dunsmoor et al., 2009). The Happy Group rated the CS+ as expressing happiness on 98% of trials, whereas the CS− was rated as expressing happiness on 2% of trials (Figure 3a). Likewise, the Fear Group rated the CS+ as expressing fear on 99% of trials, whereas the CS− was rated as expressing fear on 1% of trials (Figure 3b).

Generalization test

SCR analysis—A three-way ANOVA yielded significant main effects of Stimulus Intensity ($F(4, 144) = 11.875$, $p < .01$, $\eta_p^2 = .049$), Tertile ($F(2, 72) = 14.55$, $p < .01$, $\eta_p^2 = .023$), and Group ($F(1, 36) = 11.829$, $p < .01$, $\eta_p^2 = .159$). There was a significant Group by Stimulus Intensity interaction ($F(4, 144) = 4.77$, $p < .01$, $\eta_p^2 = .02$), as well as a significant Stimulus Intensity by Tertile interaction ($F(8, 288) = 2.463$, $p = .013$, $\eta_p^2 = .014$). The Group by Stimulus Intensity by Tertile interaction was not significant ($F(8, 288) = 1.421$, $p = .187$, $\eta_p^2 = .008$). These results suggest that, on average, participants' SCRs increased along the dimension of increasing stimulus intensity in both groups, but that generalization gradients differed between groups and over time (across tertiles; Figure 1c). Despite a non-significant three-way Group by Stimulus Intensity by Tertile interaction, we had an *a priori* hypothesis that the shape of generalization would differ between groups over time. Thus, to better discern potential differences of generalization of SCRs between groups, we performed follow-up repeated-measures ANOVAs across the entire generalization phase by group as well as mixed-effects models within each tertile by group.

Generalization to positive emotional stimuli—For the Happy Group, a separate one-way (Stimulus Intensity) repeated-measures ANOVA confirmed a significant main effect of Stimulus Intensity (happiness expression) on SCR ($F(4, 72) = 7.383$, $p < .01$, $\eta_p^2 = .291$). Post-hoc Bonferroni-corrected t-tests revealed significant differences between SH1 and SH4 ($p < .005$, hedges = .463), SH2 and SH3 ($p < .005$, hedges = .494), SH2 and SH4 ($p < .005$, hedges = .361), SH1 and SH3 ($p < .005$, hedges = .597), and a trending difference between SH1 and SH5 ($p = .012$, hedges = .312; Figure 1a). To test the shape of generalization, complementary mixed-effects models were performed. A mixed-effects model with a quadratic term significantly fit SCR better than a model with a linear term ($\chi^2(1)=27.105$, p

< .01), suggesting that the SCR generalization curve was characterized by a peak between the CS+ (SH3) and the highest intensity face (SH5). We next investigated this effect in each individual tertile. The quadratic term did not result in a better fit for the 1st tertile ($\chi^2(1)=0$, $p < 0.9972$). However, we observed significantly improved fits using a quadratic term in both the 2nd ($\chi^2(1)=20.6$, $p < .01$), and the 3rd tertiles ($\chi^2(1)=20.805$, $p < .01$; see Supplementary Table 1 for full model statistics). These results suggest that intensity-based generalization toward the happier faces was most pronounced during early testing, but that the gradient became more centered around the reinforced CS+ over the course of testing.

Generalization to negative emotional stimuli—For the Fear Group, a separate one-way repeated-measures ANOVA revealed a significant main effect of Stimulus Intensity (fear expression) on SCRs ($F(4, 72) = 9.719$, $p < .01$, $\eta_p^2 = .351$). Post-hoc Bonferroni-corrected t-tests revealed significant differences between SF1 and SF4 ($p < .005$), SF1 and SF5 ($p < .005$, hedges = .492), SF2 and SF5 ($p < .005$, hedges = .456), SF3 and SF5 ($p < .005$, hedges = .313), and a trending difference between SF2 and SF4 ($p = .031$, hedges = .305; Figure 1b). To test the shape of generalization, complementary mixed-effects models were performed. A mixed-effects model with a quadratic term did not significantly fit SCR better than a model with a linear term ($\chi^2(1)=.6865$, $p < 0.40$), which suggests that the SCR generalization curve was characterized by a linear peak shift towards the highest intensity face (SH5). We further probed this effect in each tertile, and found no significant difference between a quadratic and linear model in the 1st tertile, ($\chi^2(1)=1.71$, $p < 0.19$), 2nd tertile ($\chi^2(1)=0.36$, $p < 0.55$), or 3rd tertile ($\chi^2(1)=0.27$, $p < 0.60$) (see Supplementary Table 1 for full model statistics). That peak SCRs were exhibited at the most intensely expressive fearful face for each tertile indicates that generalization persisted along a gradient of fear intensity over the course of testing.

Reaction Times—A three-way ANOVA yielded a main effect of Stimulus Intensity ($F(1, 36) = 18.726$, $p < .01$, $\eta_p^2 = .33$). There was no main effect of group ($F(1, 36) = 2.22$, $p = .56$, $\eta_p^2 = .01$). During the generalization test, the SH1 and SH2 were rarely endorsed as expressing happiness, while SH3, SH4, and SH5 were consistently rated as expressing happiness (Figure 3a). Likewise, the SF1 and SF2 were rarely endorsed as expressing fear, while SF3, SF4, and SF5 were consistently rated as expressing fear (Figure 3b).

Retrospective CS+ identification—Following the generalization test, participants were asked to identify which face stimulus had been paired with the US during the course of the experiment (Figure 3c). The majority of participants in the Happy Group (12/19) correctly identified the SH3 as the CS+, whereas 1 participant misidentified the SH2 and 6/19 participants misidentified the SH4 as being the CS+. Interestingly, only 1 participant in the Fear Group correctly identified the SF as the CS+, whereas 12/19 participants misidentified the SF4 and 6/19 misidentified the SF5 as being the CS+. A Kruskal-Wallis H-test revealed a significant difference in retrospective CS+ identification between groups ($H(1) = 15.520$, $p < .01$). Participants were thus more likely to falsely identify a more intense stimulus as being the CS+ in the Fear Group than in the Happy Group.

DISCUSSION

The present study investigated whether fear conditioning to a stimulus containing a salient positive emotional feature would lead to a counterintuitive shift in fear generalization to a harmless but even more positive emotional stimulus. Specifically, we were interested in whether humans can base fear generalization on positive emotional features extracted from a learned threat, and how this is different than fear generalization based on an ecologically fear-relevant feature. We found biased fear generalization that transferred from a slightly expressive happy and fearful face to harmless faces of the same identity with an exaggerated emotional expression, extremely happy or fearful respectively. These findings provide evidence that humans are capable of extracting positive emotional features of a learned threat that may guide fear responses to stimuli that are even more intensely positive.

Asymmetric shifts in behavioral generalization toward stimuli that are more intense than the reinforced stimulus have been reported across species (Ghirlanda & Enquist, 2003). These generalization biases are usually tested along perceptual dimensions of increasing volume, brightness, or size, but have also been reported along a negatively valenced dimension of fear expression in human fear conditioning (Dunsmoor et al., 2009, 2011). These strongly asymmetrical shifts tend to be robust when, during initial training, the CS+ is intense and the CS− is faint (reviewed in Ghirlanda & Enquist, 2003) as this likely contributes to learning that ‘intensity’ is a key feature that discriminates reinforced from unreinforced stimuli.

Discrimination learning is likely a key factor in the present results as well, as subjects in both groups learned that emotional expression was the feature that discriminated the CS+ from the neutrally expressive CS−. Transposition to unreinforced stimuli that were even more unlike the CS− could explain the shift in maximal SCRs in both groups. This would fit with elemental models of stimulus generalization that propose generalization is determined by the associative strength of those elements shared with the CS (McLaren & Mackintosh, 2002). From this purely associative learning account, the valence of the features that discriminate threat from safety, whether positive or negative, would be irrelevant to shape the subsequent generalization gradient.

Yet, there was an important difference in the maintenance of physiological arousal to the maximally fearful versus happy face across groups that complicates a purely associative learning account. Over the course of the test, generalization to happy faces adopted a quadratic (inverted-U) shape, with maximal responding to the originally conditioned happy face. In contrast, generalization to fearful faces remained linear throughout the entire test period, with maximal responding to the most fearful face. Differential fear conditioning to a quasi-fearful face CS+ was also greater than to a quasi-happy face CS+, and a significant proportion of subjects in the Fear Group also mistakenly identified the more intensely fearful face stimulus as the actual threat. This suggests that fear-relevance is an important factor in maintaining biased fear generalization, whereas experiences that disconfirm threat expectations to positive stimuli lead to a faster reduction in conditioned arousal and more accurate retrospective memory of threat. Interestingly, this distinction in fear generalization along positive and negative stimulus dimensions is not predicted from recent work showing

selective fear conditioning and delayed extinction to positive emotional stimuli (Stussi et al., 2018a).

Some limitations of this study should be acknowledged. First, while successful discriminative fear conditioning was verified in both groups, differential SCRs (CS+ minus CS−) were greater in the Fear Group versus Happy Group. This difference is perhaps not surprising, given the literature on fear preparedness using fear-relevant stimuli versus fear-irrelevant stimuli (Ohman & Mineka, 2001). Notably, this difference did not seem to affect the peak shift of fear generalization SCRs observed in both groups during early generalization testing. However, the strength of initial learning may have contributed to the sustained fear generalization observed in the Fear Group over the course of testing. Secondly, although we chose happy faces as the positive stimulus continuum here, an open question is whether a similar peak shift of fear responses extends to other positive domains, such as attractiveness or trustworthiness (FeldmanHall et al., 2018). Future research is warranted to test the influence of positive features on various aspects of fear learning (e.g., Stussi, Pourtois, & Sander, 2018b).

These findings have implications for understanding patterns of overgeneralization in psychiatric disorders. Laboratory research is beginning to reveal patterns of overgeneralization in anxiety and trauma-related disorders (Dunsmoor & Paz, 2015). But it is important to consider that real-world stimuli associated with negative emotional experiences are complex and contain features that can be represented along multiple dimensions. As natural stimuli are rarely experienced in the exact same form twice, generalization requires transferring learning based on the quality and proportion of shared features across encounters with different stimuli. Thus, an adaptive learning system should extract those features that are most likely to portend danger in the future. Because not all negative experiences involve intrinsically dangerous stimuli (e.g., snakes or spiders) or situations (e.g., fear of heights), fear generalization in the real world may take seemingly unusual paths, leading to fear and avoidance of stimuli that are characteristically positive.

For example, in social phobia, a fear of social situations, like a party, might scale with the size of the event. A fear of dogs might extend to even extremely cute and fluffy animals that share typically positive features (e.g., floppy ears, fur, wagging tail) with a dog that acted ferociously toward an individual in the past. In the context of the present experiment, a smile, typically used to build trust, might take on a negative association if someone’s trust has been grossly violated in the past—thereby leading to a counterintuitive distrust of individuals with a great big smile. A better understanding of how behavior is influenced by the affective features extracted from a learned threat will expand the explanatory power of fear conditioning models for affective disorders.

These findings may also have implications for neuropsychological disorders marked by impaired facial recognition such as prosopagnosia, autism, and schizophrenia. The ability to identify objects, people, and places depends on generalizing prior information to novel experiences. Overgeneralization of an association between a feature and an affective outcome may interact with perceptual identification and result in perceptual impairment. While current literature supports an independent-systems framework for facial and

emotional recognition (Adolphs, Tranel, & Damasio, 2003; Darke, Cropper, & Carter, 2019), to our knowledge little research has been conducted on the effects of emotional generalization on facial recognition. Future research should investigate how emotional generalization might be implicated in neuropsychological disorders of facial perception.

CONCLUSION

In conclusion, these findings demonstrate new evidence for distinct patterns of fear generalization across divergent affective social stimuli. Specifically, following fear conditioning to a quasi-fearful or quasi-happy face, healthy adult subjects exhibited biased fear generalization toward faces of the same identity expressing an exaggerated emotional expression, regardless of valence. However, fear generalization was more persistent to fearful faces. Participants were also more likely to retrospectively misidentify an exaggerated face expressing fear as the fear conditioned stimuli. These results contribute to a growing understanding of the features that guide generalization of emotional learning and have implications for psychiatric disorders typified by overgeneralization of emotional features.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

We thank our colleague, Sam Cooper, for statistical assistance. This work was supported by National Institutes of Health Grant R00 MH106719 to J.E.D.

References

Adolphs R, Tranel D, & Damasio AR (2003). Dissociable neural systems for recognizing emotions. Brain and Cognition, 52(1), 61–69. 10.1016/S0278-2626(03)00009-5 [PubMed: 12812805]

Åhs F, Rosén J, Kastrati G, Fredrikson M, Agren T, & Lundström JN (2018, 12 1). Biological preparedness and resistance to extinction of skin conductance responses conditioned to fear relevant animal pictures: A systematic review. Neuroscience and Biobehavioral Reviews, Vol. 95, pp. 430–437. 10.1016/j.neubiorev.2018.10.017 [PubMed: 30381252]

Bates D, Mächler M, Bolker B, & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4 | Bates | Journal of Statistical Software. Journal of Statistical Software. Retrieved from https://www.jstatsoft.org/article/view/v067i01

Darke H, Cropper SJ, & Carter O (2019). A Novel Dynamic Morphed Stimuli Set to Assess Sensitivity to Identity and Emotion Attributes in Faces. Frontiers in Psychology, 10(4), 757 10.3389/fpsyg.2019.00757 [PubMed: 31024397]

Dunsmoor JE, Ahs F, Zielinski DJ, & LaBar KS (2014). Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. Neurobiology of Learning and Memory. 10.1016/j.nlm.2014.02.010

Dunsmoor JE, & LaBar KS (2013). Effects of discrimination training on fear generalization gradients and perceptual classification in humans. Behavioral Neuroscience. 10.1037/a0031933

Dunsmoor JE, Mitroff SR, & LaBar KS (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. Learning and Memory. 10.1101/lm.1431609

Dunsmoor JE, & Paz R (2015). Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. Biological Psychiatry. 10.1016/j.biopsych.2015.04.010

Dunsmoor JE, Prince SE, Murty VP, Kragel PA, & LaBar KS (2011). Neurobehavioral mechanisms of human fear generalization. NeuroImage. 10.1016/j.neuroimage.2011.01.041

Dymond S, Dunsmoor JE, Vervliet B, Roche B, & Hermans D (2015). Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. Behavior Therapy. 10.1016/j.beth.2014.10.001

Ekman P, & Friesen WV (1976). Measuring facial movement. Environmental Psychology and Nonverbal Behavior, 1(1), 56–75. 10.1007/BF01115465

Faul F, Erdfelder E, Lang AG, & Buchner A (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. 10.3758/BF03193146 [PubMed: 17695343]

FeldmanHall O, Dunsmoor JE, Tompary A, Hunter LE, Todorov A, & Phelps EA (2018). Stimulus generalization as a mechanism for learning to trust. Proceedings of the National Academy of Sciences. Retrieved from http://www.pnas.org/content/early/2018/01/26/1715227115.abstract

Ghirlanda S, & Enquist M (2003). A century of generalization. Animal Behaviour. 10.1006/anbe.2003.2174

Green SR, Kragel PA, Fecteau ME, & LaBar KS (2014). Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis. International Journal of Psychophysiology, 91(3), 186–193. 10.1016/j.ijpsycho.2013.10.015 [PubMed: 24184342]

Guttman N, & Kalish HI (1956). Discriminability and stimulus generalization. Journal of Experimental Psychology, 51(1), 79–88. 10.1037/h0046219 [PubMed: 13286444]

Honig WK, & Urcuioli PJ (1981). The legacy of Guttman and Kalish (1956): Twenty-five years of research on stimulus generalization. Journal of the Experimental Analysis of Behavior, 36(3), 405–445. 10.1901/jeab.1981.36-405 [PubMed: 16812256]

Jones Oliphant T, Peterson P, SciPy community E (2001). SciPy: Open Source Scientific Tools for Python. http://www.scipy.org. Retrieved from http://www.scipy.org/

LaBar KS, Crupain MJ, Voyvodic JT, & McCarthy G (2003). Dynamic perception of facial affect and identity in the human brain. Cerebral Cortex, 13(10), 1023–1033. 10.1093/cercor/13.10.1023 [PubMed: 12967919]

Lim SL, & Pessoa L (2008). Affective Learning Increases Sensitivity to Graded Emotional Faces. Emotion. 10.1037/1528-3542.8.1.96

McLaren IPL, & Mackintosh NJ (2002). Associative learning and elemental representation: II. Generalization and discrimination. Animal Learning and Behavior, 30(3), 177–200. 10.3758/BF03192828 [PubMed: 12391785]

Mednick SA, & Freedman JL (1960). Stimulus generalization. Psychological Bulletin, 57(3), 169–200. 10.1037/h0041650

Morey RA, Dunsmoor JE, Haswell CC, Brown VM, Vora A, Weiner J, … Szabo ST (2015). Fear learning circuitry is biased toward generalization of fear associations in posttraumatic stress disorder. Translational Psychiatry, 5(12), e700–e700. 10.1038/tp.2015.196 [PubMed: 26670285]

Neiderer S (2017). Norsk SlowTV - Hurtigruten - Minutt for Minutt - TV Show Part 05 of 30. Retrieved from https://www.youtube.com/watch?v=noOFJZh4KbU&t=60s

Ohman A, & Mineka S (2001). Fears, Phobias, and Preparedness: Toward an Evolved Module of Fear and Fear Learning. Psychological Review, 108(3), 483–522. 10.1037/0033-295X.108.3.483 [PubMed: 11488376]

Onat S, & Büchel C (2015). The neuronal basis of fear generalization in humans. Nature Neuroscience, 18(12), 1811–1818. 10.1038/nn.4166 [PubMed: 26571459]

Sander D, Grafman J, & Zalla T (2003). The Human Amygdala: An Evolved System for Relevance Detection. Reviews in the Neurosciences, Vol. 14, pp. 303–316. 10.1515/REVNEURO.2003.14.4.303 [PubMed: 14640318]

Seligman MEP (1970). ON THE GENERALITY OF THE LAWS OF LEARNING 1. In Psychological Review (Vol. 77).

Stussi Y, Ferrero A, Pourtois G, & Sander D (2019). Achievement motivation modulates Pavlovian aversive conditioning to goal-relevant stimuli. Npj Science of Learning, 4(1), 4 10.1038/s41539-019-0043-3 [PubMed: 31044087]

Stussi Y, Pourtois G, & Sander D (2018a). Enhanced Pavlovian aversive conditioning to positive emotional stimuli. Journal of Experimental Psychology: General, 147(6), 905–923. 10.1037/xge0000424 [PubMed: 29888941]

Stussi Y, Pourtois G, & Sander D (2018b). Enhanced Pavlovian aversive conditioning to positive emotional stimuli. Journal of Experimental Psychology: General, 147(6), 905–923. 10.1037/xge0000424 [PubMed: 29888941]

Vallat R (2018). Pingouin: statistics in Python. Journal of Open Source Software, 3(31), 1026 10.21105/joss.01026

Vanbrabant K, Boddez Y, Verduyn P, Mestdagh M, Hermans D, & Raes F (2015). A new approach for modeling generalization gradients: a case for hierarchical models. Frontiers in Psychology, 6(5), 1–10. 10.3389/fpsyg.2015.00652 [PubMed: 25688217]

Vervliet B, Kindt M, Vansteenwegen D, & Hermans D (2010). Fear generalization in humans: Impact of prior non-fearful experiences. Behaviour Research and Therapy. 10.1016/j.brat.2010.07.002

Highlights

- Prior work shows fear generalization is biased toward intense stimuli.
- We tested whether learned fear skews toward positively valenced stimuli.
- Participants generalized learned fear to both positive and negative stimuli.
- However, fear generalization to positive stimuli diminished rapidly over testing.

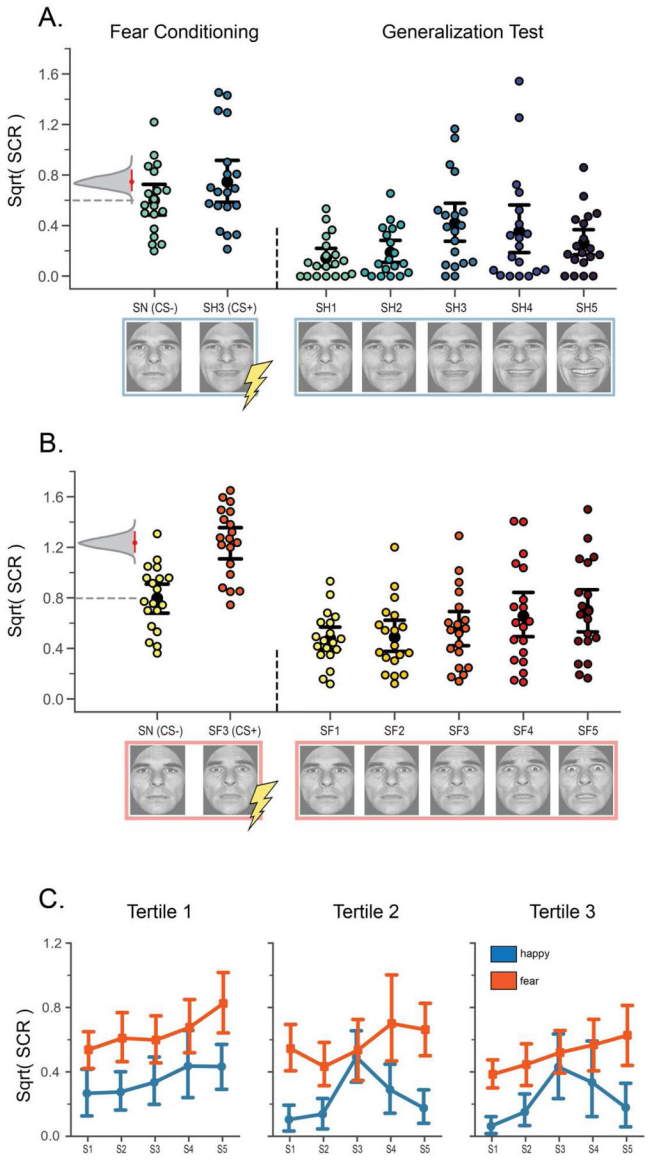

Figure 1: Skin conductance responses.

(a) Left. SCRs from fear conditioning for the Happy Group indicated successful acquisition discriminatory conditioning. Right. SCRs from the generalization test indicated generalization to stimuli of faces expressing increasing levels of happiness. (b) Left. SCRs from fear conditioning for the Fear Group indicated successful acquisition of discriminatory conditioning. SCRs from the generalization test indicated generalization to stimuli of faces expressing increasing levels of fear. (c) Generalization of SCRs split into tertiles and overlaid across groups illustrates a more persistent linear generalization for the Fear Group compared to the Happy Group. For fear conditioning, a-b Left, point and error bars correspond to the mean and 95% CI of the CS+ minus CS− difference score and is centered on the mean of the CS+, where the dashed line is the mean of the CS−. A kernel density estimate (grey) was fit to the distribution of bootstrapped difference of means. All points and error bars correspond to means with 95% CIs.

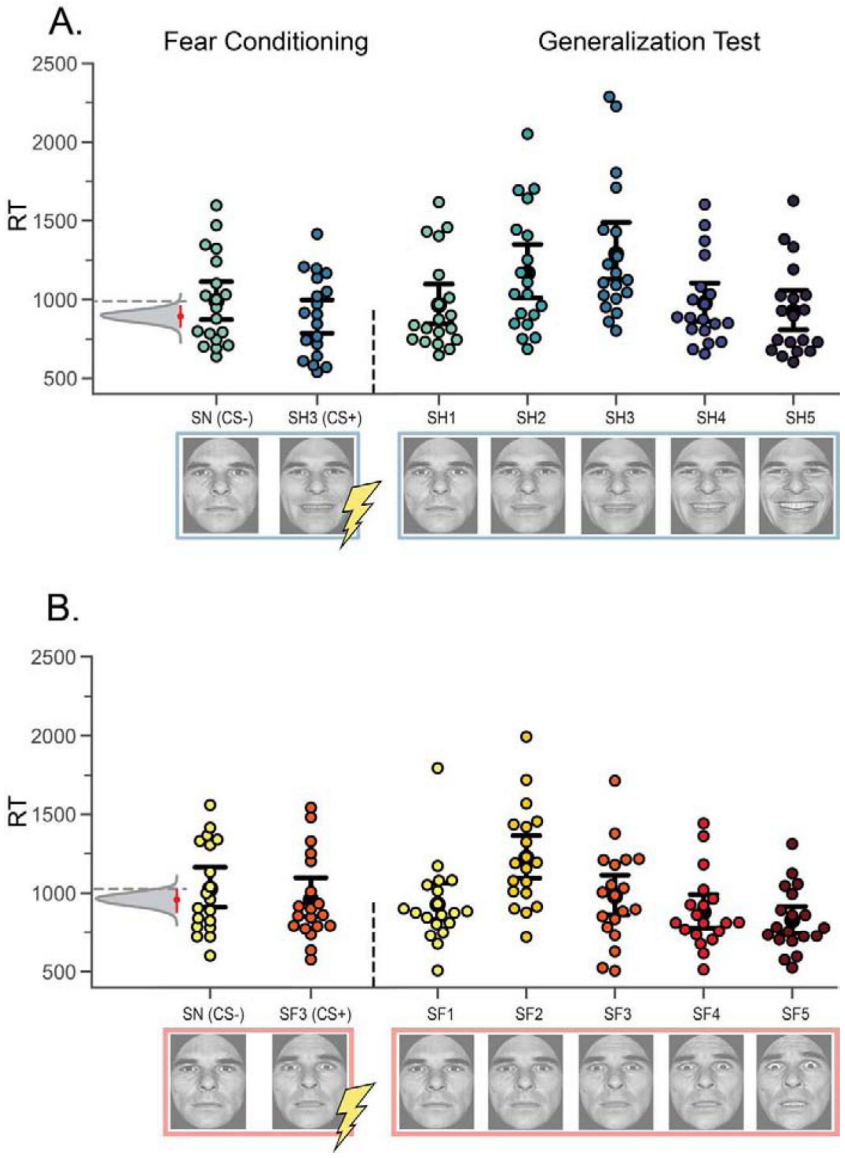

Figure 2: Reaction times.

(a) Left. RTs from fear conditioning for the Happy Group indicated faster responses to the CS+ relative to the unreinforced CS−. Right. RTs from the generalization test indicated faster responses to stimuli of faces expressing increasing levels of happiness (b) Left. RTs from fear conditioning for the Fear Group indicated faster responses to the CS+ relative to the unreinforced CS−. Right. RTs from the generalization test indicated faster responses to stimuli of faces expressing increasing levels of fear. For fear conditioning, a-b Left, point and error bars correspond to the mean and 95% CI of the CS+ minus CS− difference score and is centered on the mean of the CS+, where the dashed line is the mean of the CS−. A kernel density estimate (grey) was fit to the distribution of bootstrapped difference of means. All points and error bars correspond to means with 95% CIs.

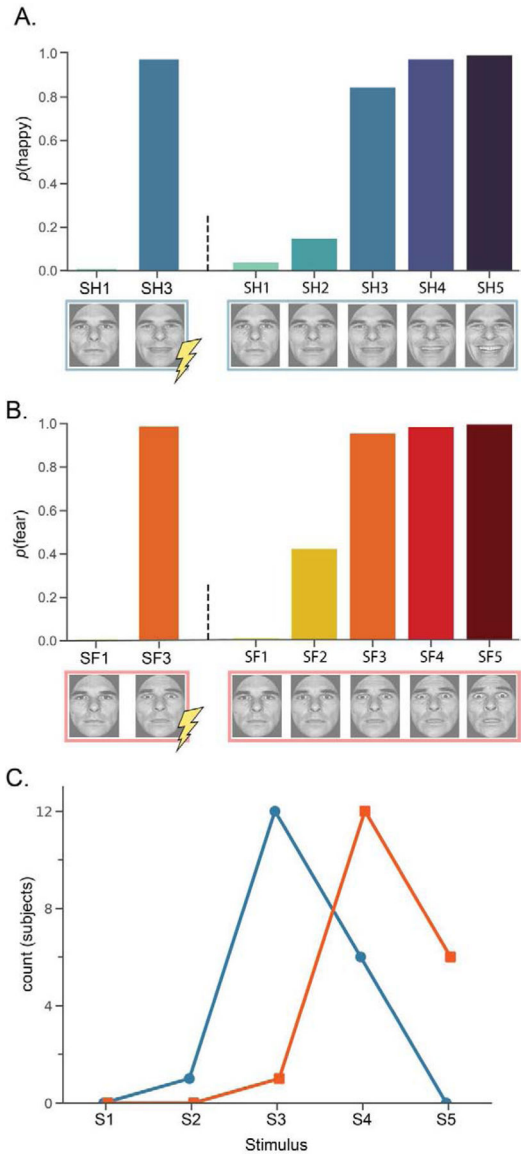

Figure 3: Subjective face rating and retrospective CS+ identification.

(a) During conditioning, participants rarely rated the CS+ as not expressing happiness. During the generalization test, participants rarely rated SH1 and SH2 as expressing happiness. (b) During conditioning, participants rarely rated the CS+ as not expressing fear. During the generalization test, participants rarely rated SH1 and SH2 as expressing fear. (c) The majority of participants in the Happy Group correctly identified the CS+ identity whereas the majority of participants in the Fear Group misidentified the CS+ identity as a face expressing stronger fear (SF4 and SF5).