This Accepted Manuscript has not been copyedited and formatted. The final version may differ from this version.

JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles: Behavioral/Cognitive

Visual Aversive Learning Compromises Sensory Discrimination

Shalev Lee[1,2], Paz Rony[3] and Avidan Galia[1,2]

[1]*Department of Psychology, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel.*
[2]*Zlotowski Center for Neuroscience, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel.*
[3]*Department of Neurobiology, Weizmann Institute of Science, Rehovot, 7610001, Israel.*

DOI: 10.1523/JNEUROSCI.0889-17.2017

Received: 29 March 2017

Revised: 20 November 2017

Accepted: 30 November 2017

Published: 8 February 2018

Author contributions: L.S., R.P., and G.A. designed research; L.S. performed research; L.S. contributed unpublished reagents/analytic tools; L.S. analyzed data; L.S., R.P., and G.A. wrote the paper.

Conflict of Interest: The authors declare no competing financial interests.

Corresponding author: Prof. Galia Avidan, Email: galiaa@bgu.ac.il, Department of Psychology, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.0889-17.2017

Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2018 the authors

JNeurosci Accepted Manuscript

JNeurosci Accepted Manuscript

Visual Aversive Learning Compromises Sensory Discrimination

Abbreviated title: Visual aversive learning decreases discrimination

Shalev Lee[1,2], Paz Rony[3], Avidan Galia[1,2*]

[1]Department of Psychology, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel.

[2]Zlotowski Center for Neuroscience, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel.

[3]Department of Neurobiology, Weizmann Institute of Science, Rehovot, 7610001, Israel.

Corresponding author:

Prof. Galia Avidan

Email: galiaa@bgu.ac.il

Department of Psychology, Ben Gurion University of the Negev, Beer Sheva, 8410501, Israel

Number of pages: 44

Number of figures: 5

Number of tables: 2

Numbers of words – Abstract: 249

Numbers of words – Introduction: 641

Numbers of words – Discussion: 1380

The authors declare no competing financial interests.

JNeurosci Accepted Manuscript

Abstract

Aversive learning is thought to modulate perceptual thresholds, which can lead to over-generalization. However, it remains undetermined if this modulation is domain specific or a general effect. Moreover, despite the unique role of the visual modality in human perception, it is unclear whether this aspect of aversive learning exists in this modality. The current study was designed to examine the effect of visual aversive outcomes on perception of basic visual and auditory features. We tested the ability of healthy participants, both males and females, to discriminate between neutral stimuli, before and after visual learning. In each experiment, neutral stimuli were associated with aversive images in an experimental group and with neutral images in a control group. Participants demonstrated a deterioration in discrimination (higher discrimination thresholds) only after aversive learning. This deterioration was measured for both auditory (tone frequency) and visual (orientation and contrast) features. The effect was replicated in five different experiments and lasted for at least 24 hours. fMRI neural responses and pupil size were also measured during learning. We showed an increase in neural activations in the anterior cingulate cortex, insula, and amygdala during aversive compared to neutral learning. Interestingly, the early visual cortex showed increased brain activity during aversive compared to neutral context trials, with identical visual information. Our findings imply the existence of a central multi-modal mechanism, which modulates early perceptual properties, following exposure to negative situations. Such a mechanism could contribute to abnormal responses that underlie anxiety states, even in new and safe environments.

Significance statement

JNeurosci Accepted Manuscript

Using a visual aversive learning paradigm, we found deteriorated discrimination abilities for visual and auditory stimuli, associated with visual aversive stimuli. We showed increased neural activations in the anterior cingulate cortex, insula, and amygdala during aversive compared to neutral learning. Importantly, similar findings were also evident in the early visual cortex during trials with aversive/neutral context, but with identical visual information. The demonstration of this phenomena in the visual modality is important, as it provides support to the notion that aversive learning can influence perception via a central mechanism, independent of input modality. Given the dominance of the visual system in human perception, our findings hold relevance to daily life, as well as imply a potential etiology for anxiety-disorders.

Introduction

A major challenge facing organisms is distinguishing between alike stimuli (discrimination), while responding similarly, when stimuli are likely related (generalization). Both abilities are crucial for survival, enabling appropriate responses to diverse situations (Guttman and Kalish, 1956; Solomon and Moore, 1975; Rescorla, 1976; Shepard, 1987; McLaren and Mackintosh, 2002; Bouton, 2006). In the case of a conditioned stimulus (CS), which predicts an aversive unconditioned stimulus (US), theory and evidence suggest a response-bias. While some studies report that aversive conditioning increases stimulus discrimination or detection (Li et al., 2008; Åhs et al., 2013), a large body of work indicates that aversive learning results in a wider generalization of the CS (Watson and Rayner, 1920; Pavlov, 1927; Hearst, 1960; Dunsmoor et al., 2009; Lissek, 2012; Vervliet et al., 2013; Dunsmoor and Paz, 2015). Wider generalization of a stimulus, paired with an aversive outcome, can facilitate a fast and efficient defensive behavior to similar stimuli. The decreased discrimination between these similar stimuli may occur already at the perceptual level (Schechtman et al., 2010; Resnik et al., 2011; Struyf et al., 2015; Zaman et

JNeurosci Accepted Manuscript

al., 2015). Evidence supporting this mechanism arises from a recent series of studies, in which participants learned to associate neutral auditory tones (the CS) with aversive odors (Resnik et al., 2011), negative sounds (Resnik et al., 2011), or monetary loss (Laufer and Paz, 2012; Laufer et al., 2016). Following conditioning, participants exhibited increased auditory thresholds and failed to discriminate new tones from the original CS.

Changes in discrimination thresholds following aversive learning have been attributed to the activity of various brain regions, including the amygdala, insula, and anterior cingulate cortex (ACC) (Laufer and Paz, 2012; Laufer et al., 2016). Alternate function of these brain regions, during aversive learning, may account for the inconsistent reports described above regarding discrimination. For example, insular activity was correlated with generalization (Laufer and Paz, 2012; Laufer et al., 2016), but also showed pattern similarity (less generalization) between the conditioned and aversive stimulus (Onat and Buchel, 2015). Furthermore, changes in tuning properties of neurons in the primate amygdala may explain how stimulus generalization and better detection can exist side by side (Resnik and Paz, 2015).

Aversive stimuli may induce plasticity in early sensory regions via a central mechanism, independent of the specific input modality. Consistent with the perceptual hypothesis, studies in rodents (Aizenberg and Geffen, 2013) and humans (Laufer et al., 2016) provide evidence for the role of the auditory cortex in the underlying plasticity. However, despite the prominent role of the visual modality in human perception, the effect of visual aversive stimuli (as US) on discrimination, particularly of neutral basic features of visual perception (as CS), has been less studied (Dunsmoor and LaBar, 2013; Struyf et al., 2015). If indeed visual aversive learning leads to alternations also in visual discrimination, it would further imply the existence of a central mechanism. Furthermore, in humans, the visual system plays a role in the development of

JNeurosci Accepted Manuscript

anxiety states. Thus, alternations in visual perception during aversive learning, may underlie anxiety disorders, which are often characterized by over-generalization (Jovanovic and Ressler, 2010; Lissek et al., 2010; Pitman et al., 2012; Lissek et al., 2014; Dunsmoor and Paz, 2015; Dymond et al., 2015; Laufer et al., 2016).

We tested the role of the visual system in over-generalization with a set of five experiments. In these experiments participants learned to associate auditory or visual neutral stimuli (CS) with aversive visual images (US). The effect of this pairing on auditory or visual discrimination thresholds was examined. In addition, we used functional imaging (fMRI) to track brain activations during learning. We hypothesized that aversive images would induce an increase in discrimination thresholds for both auditory (tone-frequency) and visual (contrast and orientation) features. We predicted that this increase will be accompanied by differences in brain activity measured during aversive visual learning, including in early visual brain regions.

Materials and Methods

Participants. A total of 315 healthy students (between 19 to 33 years of age, mean age=24.12 years, SE=0.11) from Ben-Gurion University of the Negev, Israel were recruited for the study (Experiment 1: 99 participants - 61 females, 38 males, Experiment 2: 63 participants - 49 females, 14 males, Experiment 3: 58 participants - 36 females, 22 males, Experiment 4: 59 participants, 34 females, 25 males, Experiment 5: 36 participants - 30 females, 6 males). They either received a course credit or a monetary payment to compensate for their time. Experimental procedures for the behavioral studies were approved by the ethics committee of the Psychology Department at Ben-Gurion University of the Negev and by the Helsinki committee of Soroka Medical Center, Beer-Sheva, for the fMRI experiments. All participants provided written

JNeurosci Accepted Manuscript

118 informed consent. More information regarding participants included in each experiment is
119 provided in the following relevant sections and in Table 1.

120 **General Experimental Procedure.** The framework for the experimental procedure was
121 adapted from earlier works (Resnik et al., 2011; Laufer and Paz, 2012). In all experiments, each
122 participant first performed a two alternative forced choice (2AFC) discrimination task to define
123 his/her baseline perceptual thresholds (JND – just noticeable difference) for the neutral stimuli: 1
124 kHz or 2 kHz tones in the auditory discrimination experiment (Experiment 1), low or high
125 contrast Gabors in the contrast discrimination experiment (Experiment 2), and black stripes in a
126 vertical or horizontal orientation in the orientation discrimination experiments (Experiments 3, 4,
127 5). Following the discrimination task, participants completed a conditioning (learning) session,
128 where one of the neutral stimuli was paired with either aversive or neutral images (CS+), and the
129 other stimulus with blank screens or scrambled images (CS-). Then, participants performed the
130 discrimination task again for the two neutral stimuli. In a following stage, participants also
131 performed a memory test in order to provide an indication for their attention to the stimuli during
132 the conditioning session. Finally, participants performed a validation test (image rating) for the
133 level of emotional arousal they experienced, while watching images from both categories
134 (aversive/neutral). Some of the participants in Experiments 1-3 were also re-called after 24 hours
135 and repeated the discrimination task, enabling us to estimate the stability of the effects over time.
136 In Experiment 4, we conducted the conditioning session inside a MRI scanner and recorded brain
137 activity during learning using functional imaging (fMRI). In experiment 5, we measured
138 participants' pupil size during the conditioning session and the validation test, as a physiological
139 marker for the level of emotional arousal. A schematic representation of the experimental

JNeurosci Accepted Manuscript

procedure is presented in Figure 1a. For a more detailed description of each task, see the relevant sections below and Table 2.

Discrimination Tasks. The task was a two-alternative forced choice (2AFC). In each step, two neutral stimuli were presented to the participant in a random order: the original stimulus (the CS+ or CS-) and an additional stimulus with the same physical property. These two stimuli differed only in the quantity of their physical property (x and $x+\Delta x$ respectively, where x is the original magnitude of the CS's physical property, and Δx is a small difference in this physical quantity). Participants had to decide (within a maximum duration of 10 sec), which stimulus out of these two options had the larger magnitude ($x+\Delta x$). In particular, in Experiment 1 two tones were presented (in a random order) in each trial of the discrimination task, f and $f+\Delta f$, where f is the tone frequency, and participants were asked: "Which tone had a higher pitch - first/second?". In Experiment 2 two Gabors were presented in each step, c and $c+\Delta c$, where c is the contrast of the Gabor, and we asked the participants: "Which stimuli had a higher contrast level - first/second?". In Experiment 3, 4, and 5 two black stripes were presented in different orientations, a and $a+\Delta a$, where a is the rotation angle of the stripe from a baseline orientation, and the question to the participants was: "Which stripe was more rotated clockwise - first/second?". No feedback was provided (except for a short practicing session at the beginning of the task, where feedback was presented on the screen).

The task was an adaptive two-down one-up staircase converging procedure. The magnitude of Δx (difference between the two stimuli) was decreased after two correct answers and increased after one wrong answer. Task continued until six wrong answers were obtained. The procedure converged at the stimulus level (Δx), in which the probability of a "down" response (decrease in Δx) was equal to the probability of an "up" response (increase in Δx). If x

JNeurosci Accepted Manuscript

was the original magnitude, Δx converged to the magnitude difference, at which a stimulus of x+Δx was correctly discriminated from x at 70.7% level (Levitt, 1971). Discrimination thresholds (JND) for each participant are presented as a percentage of the original stimulus magnitude, that is, Δx/x. Specific task parameters for each experiment are detailed in the Experimental Stimuli section below and in Table 2.

Conditioning Session. Two neutral stimuli (described in the General Experimental Procedure section for each experiment) were presented in this stage. One of them was assigned as CS+ and the other as CS− (30 CS+ trials and 30 CS- trials randomly counterbalanced across participants in each experiment). Each trial began with the presentation of a fixation point on the screen, followed by the presentation of the CS+ or CS- (randomly). Immediately after this (zero delay and zero overlap), an image (for CS+ stimuli) or a blank screen (for CS- stimuli) were presented. In Experiments 4 and 5, a scrambled image (the original image cut into square blocks and shuffled) was presented after the CS-. The scrambled images were used instead of the blank screens, when measuring pupil size or brain activity, to eliminate the content of the images, while maintaining a constant luminance between the two conditions. Images were aversive in the experimental group and neutral in the control group. An example of one conditioning session trial for each group in each experiment is presented in Figure 1b (Experiment 1), Figure 2a (Experiment 2), Figure 3a (Experiments 3), and Figure 4a (Experiment 4). Specific parameters for each experiment are detailed in the Experimental Stimuli section below and in Table 2.

At the end of the conditioning session, participants were inquired, whether they noticed the association between the neutral stimuli and the presentation of the images (except for Experiment 4, in which we a-priori instructed participants to be aware of the conditioning, without telling them which CS will be followed by the image). In the main analysis, we analyzed

JNeurosci Accepted Manuscript

only the results of participants, who could explicitly report that they have noticed the conditioning manipulation. This is consistent with previous studies, reporting that only participants that were aware of the conditioning (with IAPS images or electrical shocks as US) showed significant differences in conditioned reaction (skin conductance for example) to the CS (Tabbert et al., 2006; Dawson et al., 2007; Klucken et al., 2009). To strengthen the validity of our exclusion criteria, we analyzed the results of Experiment 1 also with the excluded participants. We calculated the differences between CS+ and CS- discrimination thresholds of all the participants in the experimental group (including participants from the 24-hours group). Using bootstrap analysis, we sampled (randomly with replacement) from the population of participants, which we included in the main analyses, a number of samples that was equal to the number of excluded participants. Then, we calculated the mean of those samples. This calculation was iterated 100,000 times to create a histogram of the means. We compared the distribution of this histogram to the mean of the excluded participants. The mean value for participants, who were not aware of the conditioning, was out of the 99% range of the bootstrap distribution. Thus, we used this exclusion criteria in all experiments in an unbiased way. Table 1 summarizes data about excluded participants in each experiment.

Memory Test. The memory test was conducted in order to provide an indication for participants' attention during the conditioning session. In this test, participants observed images, some of them were new and some were presented in the conditioning session. For each image, participants were required to decide, whether it was new or old. Participants with a very low performance in the memory test (less than 55% success), which calls into question their attention to the stimuli during the conditioning session, were excluded from analysis (See Table 1).

JNeurosci Accepted Manuscript

Validation Test. This test was performed in order to validate the level of emotional arousal that participants experienced, while watching the images during the conditioning session. Participants watched all the images that were presented in the conditioning session, and 10 more images selected randomly from the second category (aversive/neutral), to allow a balanced rating procedure. Participants were asked to rate each image by answering the following question: "How intensive is the emotion you feel while watching this image?" Rating was done using an analog scale of 1 (not emotional at all) to 9 (very emotional). This allowed us to compare the ratings of each participant across the two categories of images.

In the main analysis, we only analyzed results of participants, who had a difference of least 1.5 points between their ratings of the aversive and neutral categories, and that rated the aversive images with an average score that was higher than 4 (scale was 1 to 9). These values were chosen based on our cutoff criteria for image selection from the IAPS database (See the Experimental Stimuli section below). We tested the validity of this exclusion criteria by analyzing the results of Experiment 1 also with the excluded participants. We calculated the differences between CS+ and CS- discrimination thresholds of all participants in the experimental group (including participants from the 24-hours group). Then, we used bootstrap analysis, as described in the Conditioning Session section above. The mean value for participants, who did not rate the images by our criteria, was out of the 99% range of the bootstrap distribution. Hence, we used this exclusion criteria in all experiments in an unbiased way. Table 1 summarizes data about excluded participants in each experiment.

Experimental Stimuli. Experiments were conducted in a dimly lit room. Participants sat in front of a PC screen, while their head was positioned in a chinrest and their eyes were located 60 cm from the screen. The conditioning session in Experiment 4 was conducted inside a MRI

JNeurosci Accepted Manuscript

scanner and presented on a LCD screen located in the back of the scanner bore, behind the participant's head. Inside the scanner, participants viewed the stimuli through a tilted mirror mounted above their eyes on the head coil. Stimulus generation, presentation, and behavioral data acquisition and analysis were implemented in Matlab (Mathworks Inc., Natick, MA, USA). Specific experimental parameters for each experiment are described next and summarized in Table 2.

Experiment 1: The auditory conditioned stimuli were pure tones of either 1 kHz or 2 kHz (counterbalanced as CS+/CS- across participants) with a duration of 250 ms and onset/offset ramps of 5 ms, for a total of 260 ms. Tones were delivered through headphones (Philips SHL3000). In the discrimination tasks, Δf (change from original frequency) was 10% of the original tone at the beginning and was increased/decreased according to participant's performance. In the conditioning session, each trial began with the presentation of a fixation screen for 2 sec, then the tone was delivered. Following, an image for CS+ trials or a blank screen for CS- trials were presented for 200 ms. Inter trials intervals randomly ranged between 2-3 sec.

Experiment 2: The conditioned stimuli were Gabors, presented in the middle of screen (size of 50*50 mm, 128*128 pixels, visual angle of 4.77^0), at two different and easy to distinguish contrast levels (based on pilot experiments). Contrast level was calculated as $(C_{max}-C_{min})/(C_{max}+C_{min})$, with $C_{max,}$ C_{min} representing the highest and lowest luminance of each Gabor. A Gabor with high contrast of 140 and a Gabor with low contrast of 60, were used as the CS+/CS- stimuli (counterbalanced across participants). In the discrimination task, each trial included the presentation of the first Gabor for 2 sec, then a random pattern of white noise was presented (in the center of screen, within a square at the same size of the Gabor) for 0.1 sec,

JNeurosci Accepted Manuscript

254 followed by the second Gabor for 2 sec. Δc between the two Gabors in each trial started at 10%
255 from the original contrast and was modified according to performance. In the conditioning
256 session, each trial began with the presentation of a fixation screen for 2 sec, then the Gabor was
257 presented for 1 sec. Following, an image for CS+ trials or a blank screen for CS- trials were
258 presented for 200 ms. Inter trials intervals randomly ranged between 2-3 sec.

259 Experiments 3: The conditioned stimuli were black stripes, presented in the center of the
260 screen (size of 60*7 mm, 153*18 pixels, visual angle of 5.72^0), in two different orientations
261 (counterbalanced as CS+/CS- across participants). One of the stripes was in a 2^0 angle from the
262 horizontal orientation ("horizontal stripe") and the other, in a 2^0 angle from the vertical
263 orientation ("vertical stripe"). We did not use straight horizontal/vertical orientations, because
264 this could have made the orientation changes in the discrimination tasks too easy to distinguish.
265 In the discrimination tasks, each trial included the presentation of the first black stripe for 0.3
266 sec, then a white noise was presented (in the center of screen, within a square at the same size of
267 the stimuli) for 1.3 sec, followed by the second black stripe for 0.3 sec. Δa between the stripes
268 started at 5^0 and modified according to performance. In the conditioning session, each trial began
269 with the presentation of a fixation screen for 2 sec, then the stripe was presented for 1 sec.
270 Following, an image for CS+ trials or a blank screen for CS- trials were presented for 200 ms.
271 Inter trials intervals randomly ranged between 2-3 sec.

272 Experiment 4: The conditioned stimuli were black stripes counterbalanced as CS+/CS-
273 across participants, as in Experiment 3. Discrimination tasks were the same as in Experiment 3.
274 During the conditioning session of this experiment, brain activity was recorded using functional
275 imaging (fMRI). Here, we used a 50% partial conditioning: 30 non-reinforced CS+ trials and 30
276 non-reinforced CS- trials were added to the 30 CS+ and 30 CS- reinforced trials. The non-

JNeurosci Accepted Manuscript

277 reinforced trials included the presentation of the CS+ or CS- (black stripes), without pairing them
278 to the US. The non-reinforced trials were randomly interleaved between the reinforced trials, so
279 participants could not predict which stripe will be associated with an image and which will not.
280 In this manner, we could measure brain activity resulted from learning and not from the
281 presentation of the US (images). The conditioning session was divided into 3 separate event-
282 related runs. Each run consisted 10 trials of each kind (CS+, CS-, non-reinforced CS+, non-
283 reinforced CS-). Each trial of the conditioning session began with the presentation of a fixation
284 screen for 3 sec, then the stripe was presented for 2 sec. Following, an image for CS+ trials, a
285 scrambled image for CS- trials (with shuffled square blocks of 8*8 pixels), or blank screens for
286 non-reinforced CS+/CS- trials were presented for 400 ms. Inter trials intervals randomly ranged
287 between 5-6 sec. We used longer stimulus presentation durations (relatively to Experiments 1-3),
288 because longer trials were required to enable brain activity to return to its baseline values prior to
289 the beginning of each trial. Participants completed a MRI scanning session, which included a 3D
290 anatomical scan, two resting state runs, 2 amygdala localizer runs, and 3 event-related
291 conditioning session runs. Only the anatomical scan and the conditioning session scans were
292 used and analyzed in the current study.

293 Experiment 5: The conditioned stimuli were black stripes counterbalanced as CS+/CS-
294 across participants, as in Experiment 3. Discrimination tasks were the same as in Experiment 3.
295 In this experiment, pupil size was recorded during the conditioning session and the validation
296 test, in order to obtain a physiological validation for the intensity of arousal that participants
297 experienced during image watching. In the conditioning session, each trial began with the
298 presentation of a fixation screen for 3 sec, then the stripe was presented for 2 sec. Following, an
299 image for CS+ trials or a scrambled image for CS- trials (with shuffled square blocks of 8*8

JNeurosci Accepted Manuscript

pixels) were presented for 400 ms. Inter trial intervals randomly ranged between 4-5 sec. We used longer stimulus presentation durations (relatively to Experiments 1-3), because longer trials were needed to allow pupil size to return to its baseline value prior to the beginning of each trial. In order to measure pupil size during the validation test, we presented the scrambled version of each image for 3 sec at the beginning of each trial and then the original image for 6 sec (Bradley et al., 2008). After the presentation of the original image, participants were asked to rate it (See Validation Test section above). The inter trial intervals in the validation test were 4 sec.

Images: The visual US in the conditioning sessions were images from the IAPS database (International Affective Picture System) (Lang, 2008). Images were emotionally aversive for the experimental group or neutral for the control group. A total of 100 images were used in this study. Half of them were previously validated as emotionally aversive and got low scores at the valence index (lower than 4 out of 9, mean score: 1.94±0.07) and high scores at the arousal index (higher than 4 out of 9, mean score: 6.37±0.08). These images included scenes of crimes, accidents, injured body parts, etc. The other half of images were previously validated as neutral and got average-high scores at the valence index (higher than 5.5, mean score: 7.18±0.11) and low scores at the arousal index (lower than 4, mean score: 4.05±0.11). These images included scenes of landscapes, people or everyday objects (Lang, 2008). We validated that an equal number of faces and body parts appeared in each category of images. In an independent rating procedure completed at the end of the experimental session (validation test), we further verified that images were rated as aversive/neutral by our participants (See Validation Test section above). All images used in the experiment were presented at the center of gaze. The numbers of the selected images from the IAPS database are listed below.

Aversive images numbers: 2352.2, 3000, 3010, 3015, 3016, 3019, 3030, 3051, 3053,

JNeurosci Accepted Manuscript

3059, 3060, 3061, 3062, 3063, 3064, 3068, 3069, 3071, 3080, 3100, 3101, 3102, 3103, 3110, 3120, 3130, 3131, 3140, 3150, 3168, 3170, 3195, 3212, 3213, 3225, 3250, 3261, 3266, 3301, 3350, 3400, 3550, 6415, 8230, 9040, 9325, 9405, 9410, 9420, 9433.

Neutral images numbers: 1450, 1605, 1900, 2026, 2039, 2102, 2151, 2156, 2191, 2217, 2222, 2235, 2273, 2299, 2308, 2314, 2332, 2339, 2342, 2347, 2358, 2359, 2370, 2377, 2382, 2384, 2388, 2390, 2393, 2411, 2488, 2530, 2594, 2980, 5210, 5390, 5831, 5836, 7001, 7009, 7026, 7041, 7052, 7493, 7505, 7507, 7509, 7512, 7513, 9260.

MRI Setup. Participants were scanned in a 3T Philips Ingenia scanner (Amsterdam, The Netherlands) equipped with a standard head coil, located at the Soroka Medical Center, Beer Sheva, Israel. fMRI BOLD contrast was acquired using the gradient-echo echo-planner imaging sequence with parallel acquisition (SENSE: factor 2.8). Specific scanning parameters were as follows: whole brain coverage 35 slices ($3\times3\times3$ mm^3), transverse orientation, 3 mm thickness, no gap, TR = 2000 ms, TE = 35 ms, flip angle=90°, FOV = 256 × 256 and matrix size 96 × 96. High-resolution anatomical volumes were acquired with a T1-weighted 3D pulse sequence ($1\times1\times1$ mm^3, 170 slices).

fMRI Data Analysis. BrainVoyager QX software package (Brain Innovation, Maastricht, The Netherlands, Version 2.8) and Matlab (Mathworks Inc., Natick, MA, USA) were used for analysis. Preprocessing included 3D motion correction, slice time correction, filtering of low temporal frequencies (slow drift), and spatial smoothing with a Gaussian kernel of 6 mm FWHM. We analyzed each conditioning session run separately for each participant, using a whole brain general linear model (GLM), conducted on a voxel wise level. Then we grouped together the data of all runs and all participants to create a random effects group analysis GLM.

JNeurosci Accepted Manuscript

346 The four conditioning session events were used as predictors: CS+ trials, CS- trials, non-
347 reinforced CS+ trials, non-reinforced CS- trials. To define brain activity differences in aversive
348 compared to non-aversive learning, we used a contrast of (experimental group non-reinforced
349 CS+ trials – experimental group non-reinforced CS- trials) > (control group non-reinforced CS+
350 trials – control group non-reinforced CS- trials). In addition, to validate the aversive nature of
351 learning in the experimental group compared to the control group, we used a contrast of
352 (experimental group CS+ trials – experimental group CS- trials) > (control group CS+ trials –
353 control group CS- trials). We used the false discovery rate (FDR) procedure for correction of
354 multiple comparisons. Significant activity was defined at $q(FDR) < 0.05$. Anatomical brain
355 regions were identified based on known anatomical and functional landmarks according to the
356 Talairach Brain Atlas and previous studies (Laufer and Paz, 2012; Laufer et al., 2016).

357 **Eye Tracker and Pupil Size Analysis.** We recorded participants' pupil size during the
358 conditioning session and the validation test, using a video-based desktop mounted eye tracker,
359 with a sampling rate of 1000 Hz (Eye Link 1000, SR Research, Ontario, Canada). At the
360 beginning of each recording session (before the conditioning session and the validation test), the
361 system was calibrated using a display of 9 points presented in a random order on the screen. The
362 same display was also used for validation of the system. At the beginning of each trial, a fixation
363 point appeared at the center of the screen, and participants had to fixate on this point and trigger
364 the initiation of the trial by pressing a key on the keyboard (drift correction).

365 Baseline pupil size was calculated for each trial, as the average pupil size during the last
366 200 ms of fixation screen presentation (immediately before the presentation of the first stimulus).
367 In order to normalize pupil size values, the average baseline was subtracted from all pupil size
368 samples. Empty samples due to blinking were identified, and linear interpolation was used to

JNeurosci Accepted Manuscript

estimate pupil size during these missing samples. In the conditioning sessions, average pupil size was calculated (for both scrambled and original images) as the mean pupil size value in a window of 2 to 6 sec after image onset (to avoid the light reflex) (Bradley et al., 2008). Images were no longer shown on the screen during that period, but the effect on pupil size carried on after the stimuli had disappeared. In the validation tests, average pupil size for the scrambled images was calculated as the mean pupil size in a window of 2 to 3 sec following image onset (to avoid the light reflex), and the average size for the original images was calculated as the mean value in a window of 2 to 6 sec after image onset. A technical problem that occurred while recording data during the conditioning sessions of 4 participants (1 in control group and 3 in experimental group), precluded data collection and so the available partial recordings for these participants were not used for our calculations.

Experimental Design and Statistical Analysis. Statistical tests for the behavioral experiments were conducted using a two-way repeated measures ANOVA test, with two levels of group: experimental/control and two levels of condition: CS+/CS-. Post-hoc tests were used for examining specific contrasts with a-priori hypotheses. Paired t-tests were used to evaluate within-group JND changes from baseline thresholds. Imaging data in Experiment 4 were analyzed using a random effects group analysis GLM, with a false discovery rate (FDR) procedure for correction of multiple comparisons (see fMRI Data Analysis section for more details). Pupil size changes in Experiment 5 were compared using an unpaired t-test (between-group analysis) in the conditioning session or a two-way ANOVA test (with two levels of group: experimental/control and two levels of image valence: aversive/neutral) in the validation test. Numbers of participants used for analysis in each experiment are summarized in Table 1. Statistical analyses were conducted using Statistica (Dell Inc., TX, USA, version 13), Matlab

JNeurosci Accepted Manuscript

(Mathworks Inc., Natick, MA, USA), and BrainVoyager QX software package (Brain Innovation, Maastricht, The Netherlands, Version 2.8). Results are presented as mean±SE.

Results

To examine the effect of visual aversive learning on neutral stimuli discrimination thresholds (JND, just noticeable differences), we conducted a classical conditioning session. In this session participants learned to associate neutral stimuli (conditioned stimuli, CS) with the appearance (CS+) or absence (CS-) of an image (unconditioned stimuli, US). Two categories of images were selected from the IAPS database (Lang, 2008): aversive (for the experimental group) and neutral (for the control group). In order to validate participants' emotional responses, we asked them to rate the intensity of their emotion, while watching each of the conditioning session images (validation test). Rating was done on an analog scale of 1 (not emotional at all) to 9 (very emotional). Across all participants in all experimental groups tested in the present study (n=301, after exclusion of 14 participants, who did not perform well in the discrimination and memory tests, see Materials and Methods section and Table 1 for details), aversive images were rated as significantly more emotional than neutral images (paired t-test, $t(300)=30.17$, $p=0.000$, average score for aversive images: 6.85±0.09, for neutral images: 3.09±0.09). In a separate experiment (Experiment 5), we used pupil size as a physiological marker of fear, to further validate the intensity of arousal participants experienced, while watching the neutral and aversive images (Steinhauer et al., 2004; Bradley et al., 2008; Tavakoli et al., 2014; R.-Tavakoli et al., 2015). As expected, we found a larger change in pupil size for aversive compared to neutral images (see Experiment 5 at the end of this section for more details).

To detect changes in discrimination thresholds (JND), which result from aversive learning, we evaluated participants' JND values around each of the neutral stimuli. This was

JNeurosci Accepted Manuscript

done using a two-alternative forced choice (2AFC) task (see Materials and Methods section for more details). This discrimination task was conducted before and immediately after the conditioning session. At the end of the conditioning session of each experiment, participants were asked whether they noticed the association between the CS and the presentation of the images. We analyzed only the results of participants, who could explicitly report that they have noticed the conditioning manipulation (consistent with previous findings (Tabbert et al., 2006; Dawson et al., 2007; Klucken et al., 2009)), and who rated the aversive images as more aversive than the neutral images. See Materials and Methods section for more details regarding these exclusion criteria and their validation. Information regarding inclusion and exclusion of participants in each experiment is summarized in Table 1. A general description of the experimental protocol is presented in Figure 1a.

Effect of visual aversive conditioning on auditory discrimination (Experiment 1)

In this first experiment, auditory neutral sounds were used as the CS in the conditioning session. One out of 1 kHz or 2 kHz pure tones (CS+, counterbalanced) was paired with images, and the other tone (CS-) with blank screens (Figure 1b). There was no difference in the effect of aversive stimuli on JND values of the 1 kHz (n=11) and 2 kHz (n=13) tones as CS+ (unpaired t-test, $t(22)=0.27$, $p=0.78$). Analyses of JND values for each condition in each of the two groups (experimental group n=24, control group n=24), revealed that participants showed a significant decrease in threshold (improvement in discrimination) following conditioning to the CS- tone (unpaired with images), compared to baseline threshold measured before conditioning. There was a decrease of -27.81±12.26% from baseline threshold in the experimental group (paired t-test, $t(23)=2.27$, $p=0.03$) and a decrease of -41.5±9.37% in the control group (paired t-test, $t(23)=4.43$, $p=0.0002$). For the CS+ tone in the control group (paired with neutral images) we

JNeurosci Accepted Manuscript

438 also found an improvement in performance, and a decrease of -45.39±8.54% from baseline
439 (paired t-test, $t(23)=5.31$, $p=0.00002$). These findings are in accordance with previous studies
440 demonstrating that mere repeated exposure to auditory stimuli can improve performance
441 (Ahissar and Hochstein, 1996; Amitay et al., 2006; Ortiz and Wright, 2009). Further, there was
442 no significant difference between the three cases (repeated-measures ANOVA, CS+ versus CS-
443 for control group: $F(1,46)=0.046$, $p=0.83$, experimental versus control group for CS-:
444 $F(1,46)=0.786$, $p=0.38$).

445 In contrast, experimental group participants did not improve in their performance for the
446 CS+ tone (paired with aversive images). An increase of 32.1±27.65% in their JND, compared to
447 pre-conditioning baseline, was observed. This was significantly higher from performance of the
448 same participants around the CS- tone and from performance of control participants around the
449 CS+ tone, with an interaction effect between the CS+ and CS- across the experimental and
450 control groups (repeated-measures ANOVA, CS*group interaction: $F(1,46)=6.17$, $p=0.017$,
451 $\eta^2=0.12$, CS+ versus CS- for experimental group: $F(1,46)=10.88$, $p=0.002$, experimental versus
452 control group for CS+: $F(1,46)=7.17$, $p=0.01$). Thus, participants deteriorated in their
453 performance around tone frequencies, which were conditioned to visual aversive stimuli,
454 compared to tones, which were paired with neutral stimuli (Figure 1c). The effect was observed
455 for the CS+ tone in 45.83% of the participants in the experimental group compared to 4.17% of
456 the participants in the control group (Fisher's exact test between groups: $p=0.0009$) (Figure 1d).

457 We next asked if the perceptual changes are maintained overnight, and can therefore
458 point to perceptual learning rather than short-term adaptation. To do so, we measured perceptual
459 thresholds in a new group of participants (n=19). In this experiment, we only used one
460 experimental group with aversive conditioning session. Participants were re-called after 24 hours

JNeurosci Accepted Manuscript

461 to perform the threshold discrimination test one more time. The results of the original experiment
462 were replicated with this new group (paired t-test, $t(18)=2.65$, $p=0.02$), as well as remained
463 stable for at least 24 hours (paired t-test, $t(18)=2.14$, $p=0.046$) (Figure 1e).

464 These results revealed that conditioning to aversive images increased discrimination
465 thresholds of auditory neutral stimuli. The finding is consistent with previous studies, showing
466 the same effect using conditioning to aversive odors and sounds (Resnik et al., 2011). In the
467 following experiments, we examined whether this effect could also be found in discrimination
468 tests of visual neutral stimuli paired with unconditioned visual images.

469 **Effect of visual aversive conditioning on contrast discrimination (Experiment 2)**

470 In this experiment, we used Gabors with two different contrast levels (high and low), as
471 conditioned visual stimuli (experimental group n=29, control group n=19) (Figure 2a).
472 Experimental procedure was otherwise similar to the one described for Experiment 1. There was
473 no difference in the effect of aversive stimuli on JND values of high (n=13) or low (n=16)
474 contrast Gabors as the CS+ (unpaired t-test, $t(27)=0.68$, $p=0.49$). In both groups, no change in
475 contrast discrimination threshold was observed for the Gabors of the CS- condition, when we
476 compared performance following conditioning to baseline performance (paired t-test,
477 experimental: $t(28)=1.03$, $p=0.31$, control: $t(18)=0.04$, $p=0.97$). No change was observed either
478 for the CS+ Gabor of the control group (paired t-test, $t(18)=1.02$, $p=0.32$). There was no
479 significant difference between these three results (repeated-measures ANOVA, CS+ versus CS-
480 for control group: $F(1,46)=0.52$, $p=0.47$, experimental versus control group for CS-:
481 $F(1,46)=0.22$, $p=0.64$). This is in accordance with previous studies, employing perceptual
482 learning tasks with discrimination of Gabor contrast, that did not find an improvement in
483 performance under such conditions (for example (Adini et al., 2004)). Improved performance in

JNeurosci Accepted Manuscript

visual perception tasks was reported in the literature only when a more intensive training was employed, compared to the procedure used in the present study (for example (Karni and Sagi, 1993)).

In contrast, for the CS+ Gabor of the experimental group (paired with aversive images) participants showed an increase of 42.9±13.44% in discrimination threshold, compared to their baseline JND (paired t-test, $t(28)=3.19$, $p=0.003$). This increase was significant also when compared directly to the CS- Gabor in the same group or to the CS+ Gabor in the control group, with an interaction effect between the CS+ and CS- across the experimental and control groups (repeated-measures ANOVA, CS*group interaction: $F(1,46)=7.52$, $p=0.009$, $\eta^2=0.14$, CS+ versus CS- for experimental group: $F(1,46)=12.02$, $p=0.001$, experimental versus control group for CS+: $F(1,46)=8.14$, $p=0.006$) (Figure 2b). The majority of participants in the experimental group exhibited the increase for the CS+ (72.41%) compared to the CS- (44.83%, Fisher's exact test for experimental group between Gabors: $p=0.03$), and compared to CS+ in the control group (36.84%, Fisher's exact test for CS+ Gabor between groups: $p=0.01$) (Figure 2c). Here again, most of participants in the experimental group (n=20) were re-called the following day, in order to perform the discrimination task again. The effect remained stable after these 24 hours (paired t-test, $t(19)=2.09$, $p=0.049$) (Figure 2d).

Effect of visual aversive conditioning on orientation discrimination (Experiment 3)

The goal of Experiment 3 was to examine whether the finding documented in Experiment 2 is restricted to the contrast feature, or is robust to other basic features of visual perception. We therefore employed the same paradigm as in Experiment 2, except for the use of black stripes in two different orientations, vertical and horizontal, as the CS+/CS- stimuli (experimental group n=25, control group n=22) (Figure 3a). There was no difference in the effect of aversive stimuli

JNeurosci Accepted Manuscript

on JND values, when using the horizontal (n=13) or vertical (n=12) stripes as the CS+ (unpaired t-test, $t(23)=0.31$, $p=0.76$). As in Experiment 2, there was no change in angle (orientation) discrimination thresholds compared to baseline thresholds for the CS- in both groups (paired t-test, experimental: $t(24)=0.72$, $p=0.48$, control: $t(21)=0.82$, $p=0.42$). No change was found either for the CS+ in the control group (paired t-test, $t(21)=1.26$, $p=0.22$). Additionally, there was no significant difference between these three cases (repeated-measures ANOVA, CS+ versus CS- for control group: $F(1,46)=0.0002$, $p=0.99$, experimental versus control group for CS-: $F(1,46)=0.08$, $p=0.77$).

In contrast, yet consistent with the previous experiments, participants showed an increase of 65.72±31.47% in JND values for the CS+ that was paired with aversive images in the experimental group (paired t-test, $t(24)=2.09$, $p=0.047$). This was significantly different when compared to the CS- condition in the same group or when compared to the CS+ condition in the control group (but here the CS*group interaction effect was not significant) (repeated-measures ANOVA, CS*group interaction: $F(1,45)=7.52$, $p=0.094$, $\eta^2=0.061$, CS+ versus CS- for experimental group: $F(1,46)=6.32$, $p=0.02$, experimental versus control group for CS+: $F(1,46)=5.37$, $p=0.02$) (Figure 3b). The effect occurred in the majority of experimental group participants (52%) for the CS+ condition compared to performance in the control group for the same condition (22.73%, Fisher's exact test between groups: $p=0.04$) (Figure 3c). Experimental group participants (n=23) performed the discrimination task again on the following day, and discrimination thresholds remained high (paired t-test, $t(22)=2.26$, $p=0.03$) (Figure 3d).

Together, the described experiments show that the increased discrimination thresholds following conditioning to aversive visual stimuli, are robust to the modality of the neutral stimulus (auditory/visual), and at least to some of its basic features.

JNeurosci Accepted Manuscript

Brain activity is modulated by visual aversive learning (Experiment 4)

To identify the underlying brain circuits that contribute to the observed changes in discrimination thresholds, neural activations during aversive visual learning were measured using fMRI. These activations were compared to activity during non-aversive visual learning (experimental group n=30, control group n=29). All conditioning sessions of this experiment were conducted in the course of a functional magnetic resonance imaging (fMRI) scan. The design of Experiment 4 was very similar to the one used in Experiment 3, except for the presentation of scrambled images instead of blank screens after the CS- stimuli, and the employment of a random partial conditioning of 50%. This was done by adding to the reinforced CS+ and CS- trials, a similar number of non-reinforced trials (CS+ stripe or CS- stripe paired with a blank screen) (Figure 4a). This design allowed us to measure brain activity that is purely driven by the neutral CS, as it acquires value, without the additional response to the presentation of the US (images).

The behavioral results of the discrimination task were consistent with those obtained in the previous experiments. There was no difference in the effect of aversive stimuli on JND values of the horizontal (n=16) or vertical (n=14) stripes as the CS+ (unpaired t-test, $t(28)=0.22$, $p=0.83$). For the CS- condition in both groups, there was no change in angle discrimination thresholds compared to baseline (paired t-test, experimental: $t(29)=1.14$, $p=0.26$, control: $t(28)=0.08$, $p=0.94$). Additionally, no change was found for the CS+ in the control group (paired t-test, $t(28)=0.02$, $p=0.98$). There was no difference between these three cases (repeated-measures ANOVA, CS+ versus CS- for control group: $F(1,58)=0.02$, $p=0.88$, experimental versus control group for CS-: $F(1,58)=0.11$, $p=0.74$). For the CS+ of the experimental group (paired with aversive images) participants exhibited an increase of 70.6±30.45% in threshold

JNeurosci Accepted Manuscript

compared to baseline (paired t-test, $t(29)=2.32$, $p=0.03$). This was significantly different when compared to the CS- condition in the same group or to the CS+ in the control group, with an interaction effect between the CS+ and CS- across the experimental and control groups (repeated-measures ANOVA, CS*group interaction: $F(1,57)=4.09$, $p=0.048$, $\eta^2=0.067$, CS+ versus CS- for experimental group: $F(1,57)=8.67$, $p=0.004$, experimental versus control group for CS+: $F(1,57)=4.28$, $p=0.04$) (Figure 4b).

We conducted a whole brain analysis of brain activity during the conditioning sessions, and compared the activation during aversive learning (experimental group) and non-aversive learning (control group). In a GLM group analysis, we first looked for differences in activity associated with non-reinforced CS+ and non-reinforced CS- trials (i.e. CS+ and CS- trials that were not followed by images). Notice that the visual information in those trials is essentially identical, because CS+ and CS- orientations were counterbalanced across participants. We compared the difference in activity between non-reinforced CS+ and non-reinforced CS- trials in the experimental and control groups. That is, we used a contrast of (experimental group non-reinforced CS+ trials – experimental group non-reinforced CS- trials) > (control group non-reinforced CS+ trials – control group non-reinforced CS- trials). This contrast revealed more activity in the ACC, insula, and interestingly- the early visual cortex (Figure 4c).

For validation of the aversive nature of learning in the experimental group, we also compared differences in activity during the reinforced CS+ and CS- trials (i.e. CS+ followed by aversive or neutral images and CS- followed by scrambled images). We used a contrast of (experimental group CS+ trials – experimental group CS- trials) > (control group CS+ trials – control group CS- trials). This contrast revealed more activity in the amygdala, insula and ventral occipital temporal cortex (vOTC) (Figure 4d). The results are in line with previous findings

JNeurosci Accepted Manuscript

regarding the role of the amygdala, the insula, and the ACC in fear learning (Buchel et al., 1998; Pine et al., 2001; Carter et al., 2006; Nitschke et al., 2006; Dunsmoor et al., 2007; Schiller et al., 2008; Klucken et al., 2009; Sehlmeyer et al., 2009; Laufer and Paz, 2012; Resnik and Paz, 2015; Laufer et al., 2016). In addition, the results show that the ACC, and interestingly, the early visual cortex, are involved in linking value to a previously neutral stimulus. Finally, we could not find any significant correlations between fMRI BOLD activity of the reported brain regions and JND values in the discrimination tasks.

Validation of emotional arousal using pupil size (Experiment 5):

We monitored participants' pupil size in order to obtain a physiological validation for the intensity of arousal participants experienced, while watching the images used in Experiments 1-4. Measurements were conducted, using an eye tracker, during the conditioning session and the validation test. Experiment 5 was very similar to Experiment 3, except for the presentation of scrambled images instead of blank screens after the CS- stimuli, as a reference for pupil size at similar luminance conditions. Importantly, the results of the discrimination task were replicated here (experimental group n=17, control group n=17). There was no difference in the effect of aversive stimuli on JND values of the horizontal (n=9) or vertical (n=8) stripes as the CS+ (unpaired t-test, $t(15)=1.32$, $p=0.21$). As in the previous experiments, for the CS- condition in both groups, there was no change in angle discrimination thresholds compared to baseline (paired t-test, experimental: $t(16)=0.08$, $p=0.93$, control: $t(16)=0.33$, $p=0.75$). No change was found either for the CS+ in the control group (paired t-test, $t(16)=0.71$, $p=0.49$). There was no difference between these three cases (repeated-measures ANOVA, CS+ versus CS- for control group: $F(1,46)=0.068$, $p=0.79$, experimental versus control group for CS-: $F(1,46)=0.07$, $p=0.79$). For the CS+ in the experimental group (paired with aversive images), participants

JNeurosci Accepted Manuscript

showed an increase of 83.84±33.13% in threshold compared to baseline (paired t-test, $t(16)=2.53$, $p=0.02$). This was significantly different when compared to the CS- condition in the same group or to the CS+ condition in the control group, with an interaction effect between the CS+ and CS- across the experimental and control groups (repeated-measures ANOVA, CS*group interaction: $F(1,32)=6.33$, $p=0.02$, $\eta^2=0.16$, CS+ versus CS- for experimental group: $F(1,32)=10.87$, $p=0.002$, experimental versus control group for CS+: $F(1,32)=6.7$, $p=0.01$) (Figure 5a). The effect occurred in the majority of experimental group participants for the CS+ (70.59%), compared to performance in the control group for the CS+ (29.41%, Fisher's exact test between groups: $p=0.02$) (Figure 5b).

We calculated the change in participants' pupil size (original image minus scrambled image), while watching aversive and neutral images. This change in pupil size was used as an indicator for emotional arousal. In the conditioning session, each participant was exposed to only one category of images, aversive in the experimental group (full-length recordings of pupil size: n=14) and neutral in the control group (full-length recordings of pupil size: n=16). We found a larger change in pupil size for aversive images compared to neutral images (unpaired t-test, $t(28)=2.06$, $p=0.048$). In the validation test, each participant was exposed to images from both categories, so we could compare changes in pupil size within-participants (full-length recordings of pupil size: n=17 in each group). We first verified that there was no effect for group type (repeated-measures ANOVA, $F(1,32)=0.017$, $p=0.89$). Consistent with the results of the conditioning session, changes in pupil size for aversive images were larger compared to changes for neutral images (repeated-measures ANOVA, $F(1,32)=32.68$, $p=0.000002$) (Figure 5c). The procedure used to calculate changes in pupil size is demonstrated in Figure 5d. This figure presents pupil size average normalized time course of experimental group participants during the

JNeurosci Accepted Manuscript

validation test. An example of a non-normalized time course of one experimental group participant is presented in Figure 5e.

Discussion

In the present study, we found that visual aversive conditioning increases discrimination thresholds for basic features of auditory and visual stimuli. Participants in the experimental groups deteriorated in their discrimination performance compared to their baseline abilities and compared to controls. Participants demonstrated the change in thresholds in a safe-context, during a post-learning session, and even 24-hours later. This suggests that the change in thresholds is due to perceptual learning and circuit plasticity. Moreover, we observed differential brain activations during the learning process, even when the conditioned stimulus was not followed by the aversive stimulus. The modified activity was measured in the ACC, insula, amygdala, and in the early visual cortex. Our findings demonstrate that changes in perceptual thresholds occur also in the visual domain, the main modality used by humans. The results further imply the existence of a central mechanism, which may be employed during learning to modulate and alter early sensory neural representations. Below we discuss the implications of the results.

Visual aversive learning increases discrimination thresholds of auditory neutral stimuli

Previous studies have shown that aversive outcome of different modalities can induce an increase in discrimination thresholds for tone-frequencies (Resnik et al., 2011). This in turn can contribute to the wider generalization observed following aversive conditioning (Schechtman et al., 2010; Laufer and Paz, 2012). The first experiment we describe here provides evidence that complex visual images can induce a similar effect – an increase in thresholds of tone-discrimination. The importance of this extension is double-fold. First, combined with the

JNeurosci Accepted Manuscript

aforementioned studies that used odors, sounds or monetary loss, and with other studies using pain and sensation (Struyf et al., 2015), the replication to the visual domain reinforces the notion that the mechanism is not specific to any particular modality of the outcome. Rather, it supports the assumption that this is a fundamental learning principle related to negative valence. Second, given the dominance of the visual system in human perception, and the fact that many scenarios are experienced by vision, the finding holds relevance to daily life, as well as suggests a potential mechanism for anxiety disorders (Lissek, 2012; Pitman et al., 2012; Dunsmoor and Paz, 2015; Laufer et al., 2016).

Visual aversive learning increases discrimination thresholds of visual neutral stimuli

Importantly, we show that aversive scenarios do not only increase thresholds in the auditory domain, but also in the visual domain itself. This was shown with two different basic properties of visual perception – orientation and contrast (Sagi, 2011). One recent study demonstrated wider generalization to color (wavelength) following pairing with shocks (Dunsmoor and LaBar, 2013). Yet, in that study, it is harder to conclude if there was a change in choice-bias or in perceptual thresholds, because of the continuous reinforcement of the conditioned stimuli during the generalization test and the use of a decision-based task. In our study, the just-noticeable-differences (JND) were measured with a 'bias-free' 2AFC task (Green and Swets, 1989), and in a safe-context, when the conditioned stimuli were no longer reinforced. It remains an intriguing open question, how the long term change in simple feature perception, affects the perception and generalization of complex scenarios (Dunsmoor and Paz, 2015). Studies have identified altered context generalization following fear learning (Maren et al., 2013). Therefore, it is reasonable to assume that changes in discrimination thresholds, as described here, can directly contribute to these processes, potentially via mechanisms such as

JNeurosci Accepted Manuscript

pattern-completion/separation (Donaldson and Hen, 2015). Future studies will need to address this matter. Additionally, the current study focuses on participants, who were aware of the learning procedure, and who rated the aversive images as negative. An interesting direction for a future study could be to investigate how participants' awareness and image ratings affect discrimination.

Increased activity of the amygdala, insula, ACC, and early visual cortex during aversive learning

Since the deterioration in discrimination occurs for multiple outcome and input modalities, it is conceivable to assume that it is controlled by a central brain mechanism and network. Such a network should process the incoming valence and then exert its impact on sensory processing. We identified several brain regions that were more active during aversive learning compared to non-aversive learning. One of these regions is the ACC, whose activity was correlated with generalization and amygdala activity in previous studies (Laufer and Paz, 2012; Laufer et al., 2016). Another region is the insula, which was found to be involved in perceptual generalization (Laufer and Paz, 2012; Onat and Buchel, 2015).

Most interestingly, we found that the early visual cortex was more active when the same visual information predicted an aversive outcome compared to a neutral outcome. This could result from changes embedded in the level of processing of the conditioned stimuli or due to attentional effects related to the anticipation to aversive stimuli. Given the many feedback signals from higher cortical areas to the sensory cortex, it is possible that stimulus recognition involves the allocation of attentional resources to these areas. In other words, our results could indicate prioritized processing (especially as the effects requires awareness). Therefore, the activations we observed may reflect an attentional state rather than low-level plasticity. Although this is a plausible interpretation, we believe that a stimulus specific effect that lasts overnight (24 hours),

JNeurosci Accepted Manuscript

692 as we demonstrated here, is a good indication of perceptual plasticity. The current experiments
693 and techniques (human imaging) do not allow dissociating local plasticity within early sensory
694 regions from top-down attentional effects, which can also be a form of perceptual plasticity.

695 Previous studies have shown that the early visual cortex, which is thought to play a role
696 in representing low-level stimulus features, can be modulated in various behavioral contexts and
697 experiences (Ito and Gilbert, 1999; Gilbert et al., 2000; Sawtell et al., 2003; Salazar et al., 2004;
698 Shuler and Bear, 2006; Serences, 2008). Accordingly, our results further challenge the classical
699 view of the early visual cortex as a simple feature detector. Rather, they imply that aversive
700 outcomes can modulate the low-level neural representations of basic features in early sensory
701 regions. Consequently, stimuli that contain these basic features and might therefore entail the
702 aversive outcome as well, could be processed faster, leading to a lower response time. In
703 agreement with this interpretation, a previous study found that the primary auditory cortex was
704 more active during fear conditioning and generalization in anxiety patients compared to healthy
705 participants (Laufer et al., 2016).

706 A natural candidate region mediating this effect could be the amygdala. In addition to its
707 traditional role in valence-processing and anxiety, recent work revealed a correlation between its
708 cellular or neural properties and behavioral generalization following aversive conditioning
709 (Shaban et al., 2006; Ciocchi et al., 2010; Laufer and Paz, 2012; Ghosh and Chattarji, 2014). In
710 the auditory domain, this might be due to the specific role and the anatomical projections that the
711 amygdala has with the auditory system. Indeed, plasticity that correlates with auditory
712 generalization was reported in the auditory thalamus and cortices (Han et al., 2008; Aizenberg
713 and Geffen, 2013; Aizenberg et al., 2015; Laufer et al., 2016). It is possible that the amygdala

JNeurosci Accepted Manuscript

may act during conditioning to shape representations in early visual cortices as well (Pessoa, 2010).

Over generalization in anxiety disorders

Finally, recent studies have provided evidence that over-generalization plays a key role in anxiety-disorders (Lissek, 2012; Pitman et al., 2012; Dunsmoor and Paz, 2015). Most of these studies focused on choice-behavior in risky and un-safe environments. In such a scenario, it is indeed a rationale behavior to over-generalize (i.e. have a bias). Recent findings in anxiety patients in the auditory domain (Laufer et al., 2016) and our current findings in the visual domain, imply that activity during aversive conditioning may modulates neural representations. These in turn can result in a later inability to discriminate between the stimuli, even in a safe-context, and even long after learning has ended. The results are also in-line with the evidence that failures to process safety signals contribute to anxiety (Christianson et al., 2012; Jovanovic et al., 2012). Thus, the current study strengthens the notion that perception and over-generalization play a role in anxiety. Specifically, it suggests that exposure to aversive outcome in a complex real-life scene, can result in perceptual changes for basic features of multiple modalities experienced in the scene. The result can be a complex pattern of generalization, as abnormally exhibited in anxiety-disorders.

References

Adini Y, Wilkonsky A, Haspel R, Tsodyks M, Sagi D (2004) Perceptual learning in contrast discrimination: the effect of contrast uncertainty. Journal of vision 4:993-1005.

Ahissar M, Hochstein S (1996) Learning pop-out detection: specificities to stimulus characteristics. Vision research 36:3487-3500.

JNeurosci Accepted Manuscript

736 Åhs F, Miller SS, Gordon AR, Lundström JN (2013) Aversive learning increases sensory detection
737 sensitivity. Biological psychology 92:135-141.
738 Aizenberg M, Geffen MN (2013) Bidirectional effects of aversive learning on perceptual acuity are
739 mediated by the sensory cortex. Nature neuroscience 16:994-996.
740 Aizenberg M, Mwilambwe-Tshilobo L, Briguglio JJ, Natan RG, Geffen MN (2015) Bidirectional Regulation
741 of Innate and Learned Behaviors That Rely on Frequency Discrimination by Cortical Inhibitory
742 Neurons. PLoS Biol 13:e1002308.
743 Amitay S, Irwin A, Moore DR (2006) Discrimination learning induced by training with identical stimuli.
744 Nature neuroscience 9:1446-1448.
745 Bouton M (2006) Learning and behavior: a contemporary synthesis. Sunderland, MA: Sinauer Associates.
746 Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and
747 autonomic activation. Psychophysiology 45:602-607.
748 Buchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-
749 related fMRI study. Neuron 20:947-957.
750 Carter RM, O'Doherty JP, Seymour B, Koch C, Dolan RJ (2006) Contingency awareness in human aversive
751 conditioning involves the middle frontal gyrus. Neuroimage 29:1007-1012.
752 Christianson JP, Fernando AB, Kazama AM, Jovanovic T, Ostroff LE, Sangha S (2012) Inhibition of fear by
753 learned safety signals: a mini-symposium review. J Neurosci 32:14118-14124.
754 Ciocchi S, Herry C, Grenier F, Wolff SB, Letzkus JJ, Vlachos I, Ehrlich I, Sprengel R, Deisseroth K, Stadler
755 MB, Muller C, Luthi A (2010) Encoding of conditioned fear in central amygdala inhibitory circuits.
756 Nature 468:277-282.
757 Dawson ME, Rissling AJ, Schell AM, Wilcox R (2007) Under What Conditions Can Human Affective
758 Conditioning Occur Without Contingency Awareness?: Test of the Evaluative Conditioning
759 Paradigm. Emotion (Washington, DC) 7:755-766.

JNeurosci Accepted Manuscript

760 Donaldson ZR, Hen R (2015) From psychiatric disorders to animal models: a bidirectional and
761 dimensional approach. Biol Psychiatry 77:15-21.
762 Dunsmoor JE, LaBar KS (2013) Effects of discrimination training on fear generalization gradients and
763 perceptual classification in humans. Behavioral neuroscience 127:350-356.
764 Dunsmoor JE, Paz R (2015) Fear Generalization and Anxiety: Behavioral and Neural Mechanisms.
765 Biological psychiatry 78:336-343.
766 Dunsmoor JE, Bandettini PA, Knight DC (2007) Impact of continuous versus intermittent CS-UCS pairing
767 on human brain activation during Pavlovian fear conditioning. Behav Neurosci 121:635-642.
768 Dunsmoor JE, Mitroff SR, LaBar KS (2009) Generalization of conditioned fear along a dimension of
769 increasing fear intensity. Learning & memory (Cold Spring Harbor, NY) 16:460-469.
770 Dymond S, Dunsmoor JE, Vervliet B, Roche B, Hermans D (2015) Fear Generalization in Humans:
771 Systematic Review and Implications for Anxiety Disorder Research. Behavior therapy 46:561-
772 582.
773 Ghosh S, Chattarji S (2014) Neuronal encoding of the switch from specific to generalized fear. Nat
774 Neurosci. 18(1):112-120.
775 Gilbert C, Ito M, Kapadia M, Westheimer G (2000) Interactions between attention, context and learning
776 in primary visual cortex. Vision research 40:1217-1226.
777 Green D, Swets J (1989) Signal Detection Theory and Psychophysics. Los Altos Hills, CA: Peninsula Pub.
778 Guttman N, Kalish HI (1956) Discriminability and stimulus generalization. Journal of experimental
779 psychology 51:79-88.
780 Han JH, Yiu AP, Cole CJ, Hsiang HL, Neve RL, Josselyn SA (2008) Increasing CREB in the auditory thalamus
781 enhances memory and generalization of auditory conditioned fear. Learn Mem 15:443-453.
782 Hearst E (1960) Simultaneous generalization gradients for appetitive and aversive behavior. Science
783 (New York, NY) 132:1769-1770.

JNeurosci Accepted Manuscript

784 Ito M, Gilbert CD (1999) Attention modulates contextual influences in the primary visual cortex of alert
785 monkeys. Neuron 22:593-604.
786 Jovanovic T, Ressler KJ (2010) How the neurocircuitry and genetics of fear inhibition may inform our
787 understanding of PTSD. Am J Psychiatry 167:648-662.
788 Jovanovic T, Kazama A, Bachevalier J, Davis M (2012) Impaired safety signal learning may be a biomarker
789 of PTSD. Neuropharmacology 62:695-704.
790 Karni A, Sagi D (1993) The time course of learning a visual skill. Nature 365:250-252.
791 Klucken T, Kagerer S, Schweckendiek J, Tabbert K, Vaitl D, Stark R (2009) Neural, electrodermal and
792 behavioral response patterns in contingency aware and unaware subjects during a picture-
793 picture conditioning paradigm. Neuroscience 158:721-731.
794 Lang P (2008) International affective picture system (IAPS): affective ratings of pictures and instruction
795 manual. Technical Report A-8. In. University of Florida, Gainesville, FL.
796 Laufer O, Paz R (2012) Monetary loss alters perceptual thresholds and compromises future decisions via
797 amygdala and prefrontal networks. The Journal of neuroscience : the official journal of the
798 Society for Neuroscience 32:6304-6311.
799 Laufer O, Israeli D, Paz R (2016) Behavioral and Neural Mechanisms of Overgeneralization in Anxiety.
800 Current biology 26:713-722.
801 Levitt H (1971) Transformed up-down methods in psychoacoustics. The Journal of the Acoustical Society
802 of America 49(2B):467-477.
803 Li W, Howard JD, Parrish TB, Gottfried JA (2008) Aversive Learning Enhances Perceptual and Cortical
804 Discrimination of Indiscriminable Odor Cues. Science (New York, NY) 319:1842-1845.
805 Lissek S (2012) Toward an account of clinical anxiety predicated on basic, neurally mapped mechanisms
806 of pavlovian fear-learning: the case for conditioned overgeneralization. Depress Anxiety 29:257-
807 263.

JNeurosci Accepted Manuscript

808 Lissek S, Kaczkurkin AN, Rabin S, Geraci M, Pine DS, Grillon C (2014) Generalized anxiety disorder is
809 associated with overgeneralization of classically conditioned fear. Biological psychiatry 75:909-
810 915.
811 Lissek S, Rabin S, Heller RE, Lukenbaugh D, Geraci M, Pine DS, Grillon C (2010) Overgeneralization of
812 conditioned fear as a pathogenic marker of panic disorder. The American journal of psychiatry
813 167:47-55.
814 Maren S, Phan KL, Liberzon I (2013) The contextual brain: implications for fear conditioning, extinction
815 and psychopathology. Nat Rev Neurosci 14:417-428.
816 McLaren IP, Mackintosh NJ (2002) Associative learning and elemental representation: II. Generalization
817 and discrimination. Animal learning & behavior 30:177-200.
818 Nitschke JB, Sarinopoulos I, Mackiewicz KL, Schaefer HS, Davidson RJ (2006) Functional neuroanatomy of
819 aversion and its anticipation. Neuroimage 29:106-116.
820 Onat S, Buchel C (2015) The neuronal basis of fear generalization in humans. Nature neuroscience
821 18:1811-1818.
822 Ortiz JA, Wright BA (2009) Contributions of procedure and stimulus learning to early, rapid perceptual
823 improvements. Journal of experimental psychology Human perception and performance 35:188-
824 194.
825 Pavlov I (1927) Conditioned reflexes. London: Oxford UP.
826 Pessoa L (2010) Emotion and cognition and the amygdala: from "what is it?" to "what's to be done?".
827 Neuropsychologia 48:3416-3429.
828 Pine DS, Fyer A, Grun J, Phelps EA, Szeszko PR, Koda V, Li W, Ardekani B, Maguire EA, Burgess N, Bilder
829 RM (2001) Methods for developmental studies of fear conditioning circuitry. Biological
830 psychiatry 50:225-228.

JNeurosci Accepted Manuscript

831 Pitman RK, Rasmusson AM, Koenen KC, Shin LM, Orr SP, Gilbertson MW, Milad MR, Liberzon I (2012)
832 Biological studies of post-traumatic stress disorder. Nature reviews Neuroscience 13:769-787.
833 R.-Tavakoli H, Atyabi A, Rantanen A, Laukka SJ, Nefti-Meziani S, Heikkilä J (2015) Predicting the Valence
834 of a Scene from Observers' Eye Movements. PloS one 10:e0138198.
835 Rescorla RA (1976) Stimulus generalization: some predictions from a model of Pavlovian conditioning.
836 Journal of experimental psychology Animal behavior processes 2:88-96.
837 Resnik J, Paz R (2015) Fear generalization in the primate amygdala. Nature neuroscience 18:188-190.
838 Resnik J, Sobel N, Paz R (2011) Auditory aversive learning increases discrimination thresholds. Nature
839 neuroscience 14:791-796.
840 Sagi D (2011) Perceptual learning in Vision Research. Vision Res 51:1552-1566.
841 Salazar RF, Konig P, Kayser C (2004) Directed interactions between visual areas and their role in
842 processing image structure and expectancy. The European journal of neuroscience 20:1391-
843 1401.
844 Sawtell NB, Frenkel MY, Philpot BD, Nakazawa K, Tonegawa S, Bear MF (2003) NMDA receptor-
845 dependent ocular dominance plasticity in adult visual cortex. Neuron 38:977-985.
846 Schechtman E, Laufer O, Paz R (2010) Negative valence widens generalization of learning. The Journal of
847 neuroscience : the official journal of the Society for Neuroscience 30:10460-10464.
848 Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008) From fear to safety and back: reversal of fear in the
849 human brain. The Journal of neuroscience : the official journal of the Society for Neuroscience
850 28:11517-11525.
851 Sehlmeyer C, Schoning S, Zwitserlood P, Pfleiderer B, Kircher T, Arolt V, Konrad C (2009) Human fear
852 conditioning and extinction in neuroimaging: a systematic review. PloS one 4:e5865.
853 Serences JT (2008) Value-Based Modulations in Human Visual Cortex. Neuron 60:1169-1181.

JNeurosci Accepted Manuscript

854 Shaban H, Humeau Y, Herry C, Cassasus G, Shigemoto R, Ciocchi S, Barbieri S, van der Putten H,
855 Kaupmann K, Bettler B, Luthi A (2006) Generalization of amygdala LTP and conditioned fear in
856 the absence of presynaptic inhibition. Nat Neurosci 9:1028-1035.
857 Shepard RN (1987) Toward a universal law of generalization for psychological science. Science (New
858 York, NY) 237:1317-1323.
859 Shuler MG, Bear MF (2006) Reward Timing in the Primary Visual Cortex. Science (New York, NY)
860 311:1606-1609.
861 Solomon PR, Moore JW (1975) Latent inhibition and stimulus generalization of the classically
862 conditioned nictitating membrane response in rabbits (Oryctolagus cuniculus) following dorsal
863 hippocampal ablation. Journal of comparative and physiological psychology 89:1192-1203.
864 Steinhauer SR, Siegle GJ, Condray R, Pless M (2004) Sympathetic and parasympathetic innervation of
865 pupillary dilation during sustained processing. International journal of psychophysiology : official
866 journal of the International Organization of Psychophysiology 52:77-86.
867 Struyf D, Zaman J, Vervliet B, Van Diest I (2015) Perceptual discrimination in fear generalization:
868 Mechanistic and clinical implications. Neuroscience and biobehavioral reviews 59:201-207.
869 Tabbert K, Stark R, Kirsch P, Vaitl D (2006) Dissociation of neural responses and skin conductance
870 reactions during fear conditioning with and without awareness of stimulus contingencies.
871 NeuroImage 32:761-770.
872 Tavakoli HR, Yanulevskaya V, Rahtu E, Heikkilä J, Sebe N (2014) Emotional Valence Recognition, Analysis
873 of Salience and Eye Movements. In: 2014 22nd International Conference on Pattern Recognition,
874 pp 4666-4671. Stockholm.
875 Vervliet B, Baeyens F, Van den Bergh O, Hermans D (2013) Extinction, generalization, and return of fear:
876 a critical review of renewal research in humans. Biological psychology 92:51-58.
877 Watson JB, Rayner R (1920) Conditioned emotional reactions. The American psychologist 55:313-317.

JNeurosci Accepted Manuscript

878 Zaman J, Vlaeyen JW, Van Oudenhove L, Wiech K, Van Diest I (2015) Associative fear learning and
879 perceptual discrimination: a perceptual pathway in the development of chronic pain.
880 Neuroscience and biobehavioral reviews 51:118-125.

881

882

883

884

885

886 # Tables and Figure Legends:

887 **Table 1: Composition of participants in each experiment (included and excluded).** 315
888 participants performed the experiment. 255 participants were included in analysis: 172 females,
889 83 males (mean age: 24.06 years, SE=0.13). 60 participants were excluded from analysis: 37
890 females, 23 males (mean age: 24.36 years, SE=0.24). Exclusions were due to unawareness to
891 conditioning during conditioning sessions, incoherent image rating, low performance in memory
892 tests, or low performance in discrimination tasks (these criteria are detailed in the text).

Exp No.	Group	Total	Included in total	Included females	Included males	Excluded due to unaw areness to conditioning	Excluded due to image rating	Excluded due to low performance in memory tests	Excluded due to low performance in discrimination tasks
Exp1	Exp	34	24	16	8	5	2	1	2
	Ctrl	31	24	17	7	5	0	2	0
	24-hours	34	19	10	9	7	7	0	1
Exp2	Exp	38	29	24	5	6	0	2	1
	Ctrl	25	19	15	4	4	0	1	1
Exp3	Exp	33	25	16	9	4	1	1	2
	Ctrl	25	22	11	11	0	3	0	0
Exp4	Exp	30	30	15	15	0	0	0	0
	Ctrl	29	29	19	10	0	0	0	0
Exp5	Exp	18	17	14	3	0	1	0	0
	Ctrl	18	17	15	2	1	0	0	0

JNeurosci Accepted Manuscript

Table 2: Experimental parameters. Parameters of discrimination tasks and conditioning sessions are detailed for each experiment.

Exp No.	CS Type	CS duration in discrimination tasks	CS duration in conditioning session	Fixation duration in conditioning session	US duration in conditioning session	ITI duration in conditioning session	Stimulus type after CS-	Participants were informed of conditioning before experiment	Reinforcment rate	No. trials per condition
Exp1	sounds (frequency)	260 ms	260 ms	2 sec	200 ms	2-3 sec	Blank screen	No	100%	30 CS+ 30 CS-
Exp2	Gabors (contrast)	2 sec	1 sec	2 sec	200 ms	2-3 sec	Blank screen	No	100%	30 CS+ 30 CS-
Exp3	stripes (orientation)	0.3 sec	1 sec	2 sec	200 ms	2-3 sec	Blank screen	No	100%	30 CS+ 30 CS-
Exp4	stripes (orientation)	0.3 sec	2 sec	3 sec	400 ms	5-6 sec	Scrambled image	Yes	50%	reinforced: 30 CS+ 30 CS- non-reinforced: 30 CS+ 30 CS-
Exp5	stripes (orientation)	0.3 sec	2 sec	3 sec	400 ms	4-5 sec	Scrambled image	No	100%	30 CS+ 30 CS-

Figure 1: The effect of visual aversive conditioning on auditory discrimination thresholds (Experiment 1). a. A general description of the experimental protocol. **b.** A schematic

JNeurosci Accepted Manuscript

representation of one trial from the conditioning session of the auditory experiment (Experiment 1). High (2 kHz) or low (1 kHz) pure tones were counterbalanced as CS+ (followed by images) or CS- (followed by blank screens). Images were aversive in the experimental group ("Exp") and neutral in the control group ("Ctrl"). **c.** Discrimination thresholds for tone frequency were tested before and after conditioning (experimental group n=24, control group n=24). Discrimination thresholds before conditioning were normalized to 100% (red dashed line), which was the reference for post-conditioning thresholds. A decrease in threshold (improvement) was measured in both groups for the CS– tone, and in the control group for the CS+ tone. The experimental group showed a deterioration in performance for the CS+ tone (paired with aversive images), compared to its results for the CS- tone, and to the CS+ tone in the control group, with a CS*group interaction effect. **d.** More experimental group participants deteriorated in discrimination thresholds for the CS+ tone compared to control group. The gray bars show the percentage of participants for whom the post-conditioning threshold was lower than the pre-conditioning threshold (improvement), and the black bars show the fraction of participants for whom the post-conditioning threshold was higher (deterioration). **e.** Change in discrimination thresholds (compared to baseline) for the aversive CS+ (solid line) and the CS- (dashed line) of the experimental group (n=19), immediately after conditioning and 24 hours later. Thresholds persisted stable after 24 hours.

Figure 2: The effect of visual aversive conditioning on contrast discrimination thresholds (Experiment 2). a. A schematic representation of one trial from the conditioning session of the contrast discrimination experiment (Experiment 2). Paradigm was similar to the one described in Figure 1b, except for the presentation of a high or low contrast Gabor as the CS+ or CS-. **b.**

JNeurosci Accepted Manuscript

Discrimination thresholds for contrast were tested before and after the conditioning session (experimental group n=29, control group n=19). No change in thresholds was measured in both groups for the CS– Gabors, nor in control group participants for the CS+ Gabor. However, increase in threshold (deterioration in performance) was measured in the experimental group for the CS+ Gabor (paired with aversive images), with a CS*group interaction effect. **c.** The majority of participants from the experimental group showed a deterioration in discrimination after conditioning to the CS+ Gabor. This percentage of participants was higher compared to the CS- Gabor in the same group and compared to participants in the control group for the CS+ Gabor. The gray bars show the percentage of participants for whom the post-conditioning threshold was lower than the pre-conditioning threshold (improvement), and the black bars show the percentage of participants, for whom their post-conditioning threshold was higher (deterioration). **d.** Change in discrimination thresholds (compared to baseline) for the aversive CS+ (solid line) and the CS- (dashed line) stimuli of the experimental group (n=20), immediately after and 24 hours following conditioning. Thresholds persisted stable after 24 hours.

Figure 3: The effect of visual aversive conditioning on orientation discrimination thresholds (Experiment 3). a. A schematic representation of one trial from the conditioning session of the orientation discrimination experiment (Experiment 3). Paradigm was similar to the one described in Figure 1b, except for the presentation of a vertical or horizontal black stripe as the CS+ or CS-. **b.** Discrimination thresholds for orientation were tested before and after the conditioning session (experimental group n=25, control group n=22). No change in thresholds was measured in both groups for the CS– stripes, nor in control group for the CS+ stripe. However, experimental group participants deteriorated (increase in threshold) in the CS+

JNeurosci Accepted Manuscript

orientation (paired with aversive images). **c.** The majority of participants from the experimental group showed a deterioration in discrimination thresholds after conditioning to the CS+ stripe. This percentage of participants was higher compared to control group for the same stripe. The gray bars show the percentage of the participants for whom the post-conditioning threshold was lower than the pre-conditioning threshold (improvement), and the black bars show the percentage of participants for whom the post-conditioning threshold was higher (deterioration). **d.** Change in discrimination thresholds (compared to baseline) for the aversive CS+ (solid line) and the CS- (dashed line) of the experimental group (n=23), immediately after and 24 hours following conditioning. Thresholds persisted stable after 24 hours.

Figure 4: Brain activity is modulated by visual aversive learning (Experiment 4). a. A schematic representation of one trial from the conditioning session of the fMRI experiment (Experiment 4). Paradigm was similar to the one described in Figure 3a, except for a partial conditioning of 50%: half of the trials were reinforced trials (CS+ stripe paired with image or CS- stripe paired with scrambled image), and the other half were non-reinforced trials (CS+ or CS- stripes paired with blank screens). **b.** Discrimination thresholds for orientation were tested before and after the conditioning session (experimental group n=30, control group n=29). No change in thresholds was measured in both groups for the CS– stripe, nor in the control group for the CS+ stripe. The experimental group deteriorated (increase in threshold) in the CS+ orientation (paired with aversive images), with a CS*group interaction effect. **c.** A contrast of (experimental group non-reinforced CS+ trials – experimental group non-reinforced CS- trials) > (control group non-reinforced CS+ trials – control group non-reinforced CS- trials) showed higher activations in the ACC, insula, and early visual cortex in the experimental group ("Exp") compared to the control group ("Ctrl"). Activations are shown in statistical q(FDR)<0.05

JNeurosci Accepted Manuscript

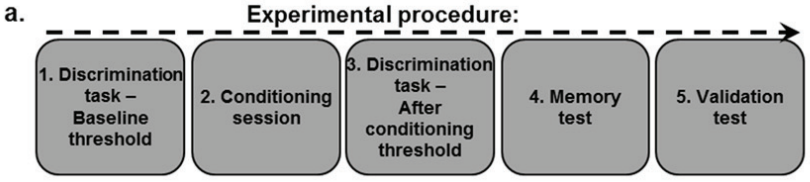
thresholds. **d.** A contrast of (experimental group CS+ trials – experimental group CS- trials) > (control group CS+ trials – control group CS- trials) showed higher activations in the amygdala, insula, and vOTC in the experimental group ("Exp") compared to the control group ("Ctrl"). Activations are shown in statistical q(FDR)<0.05 thresholds.

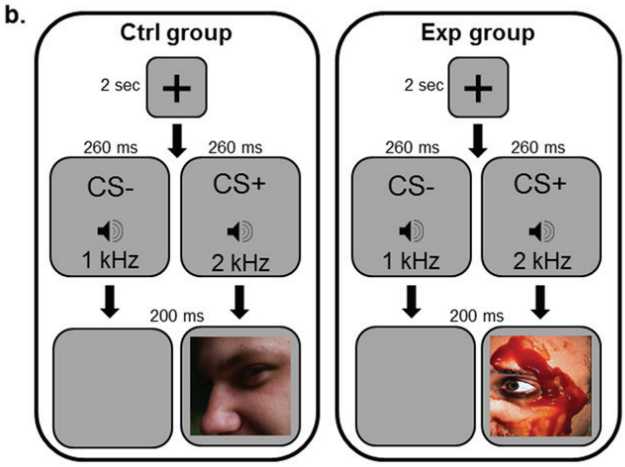
Figure 5: Validation of emotional arousal using pupil size (Experiment 5). a. Discrimination thresholds for orientation were tested before and after the conditioning session (experimental group n=17, control group n=17). No change in thresholds was measured in both groups for the CS– stripe, nor in the control group for the CS+ stripe. Experimental group deteriorated (increase in threshold) in the CS+ orientation (paired with aversive images), with a CS*group interaction effect. **b.** The majority of participants from the experimental group showed a deterioration in discrimination thresholds after conditioning to the CS+ stripe. This percentage of participants was higher compared to the control group for the same stripe. The gray bars show the percentage of participants for whom the post-conditioning threshold was lower than the pre-conditioning threshold (improvement), and the black bars show the percentage of participants, for whom their post-conditioning threshold was higher (deterioration). **c.** Change in participants' pupil size was calculated as the difference between pupil size, while watching each image and its scrambled version (see panel d for more explanations). In the conditioning session (left, between-group comparison, experimental group n=14, control group n=16), as well as in the validation test (right, within-group comparison, experimental group n=17, control group n=17), changes were larger during aversive image watching compared to neutral image watching. **d.** Demonstration of the procedure used for calculating pupil size in panel c. Average pupil size time course (normalized by baseline subtraction) of the experimental group during the validation test is

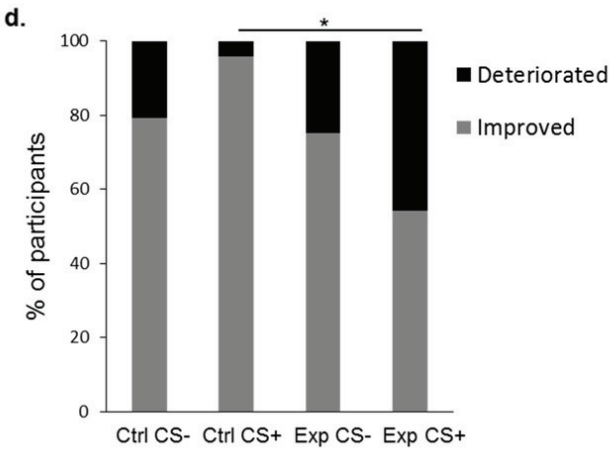
JNeurosci Accepted Manuscript

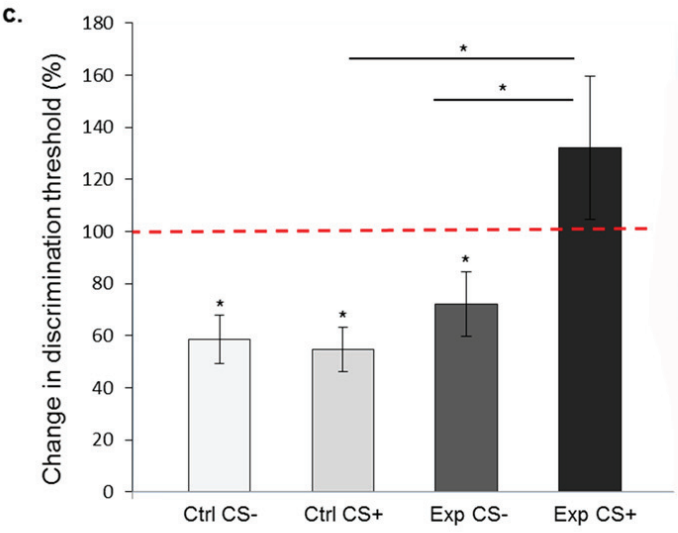
shown. The scrambled image was presented at time 0 sec (followed by pupil's light reflex), and the original image at 3 sec as shown. Average pupil sizes for the scrambled and original images were calculated as the mean value in the range marked by the red respective arrows. **e.** An example for pupil size time course (without baseline normalization) of one experimental group participant during the validation test.

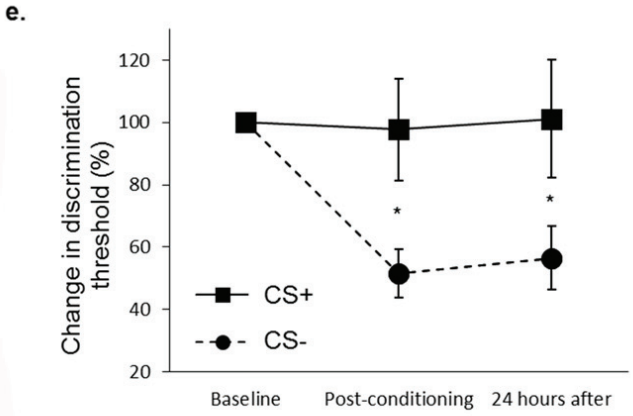
JNeurosci Accepted Manuscript



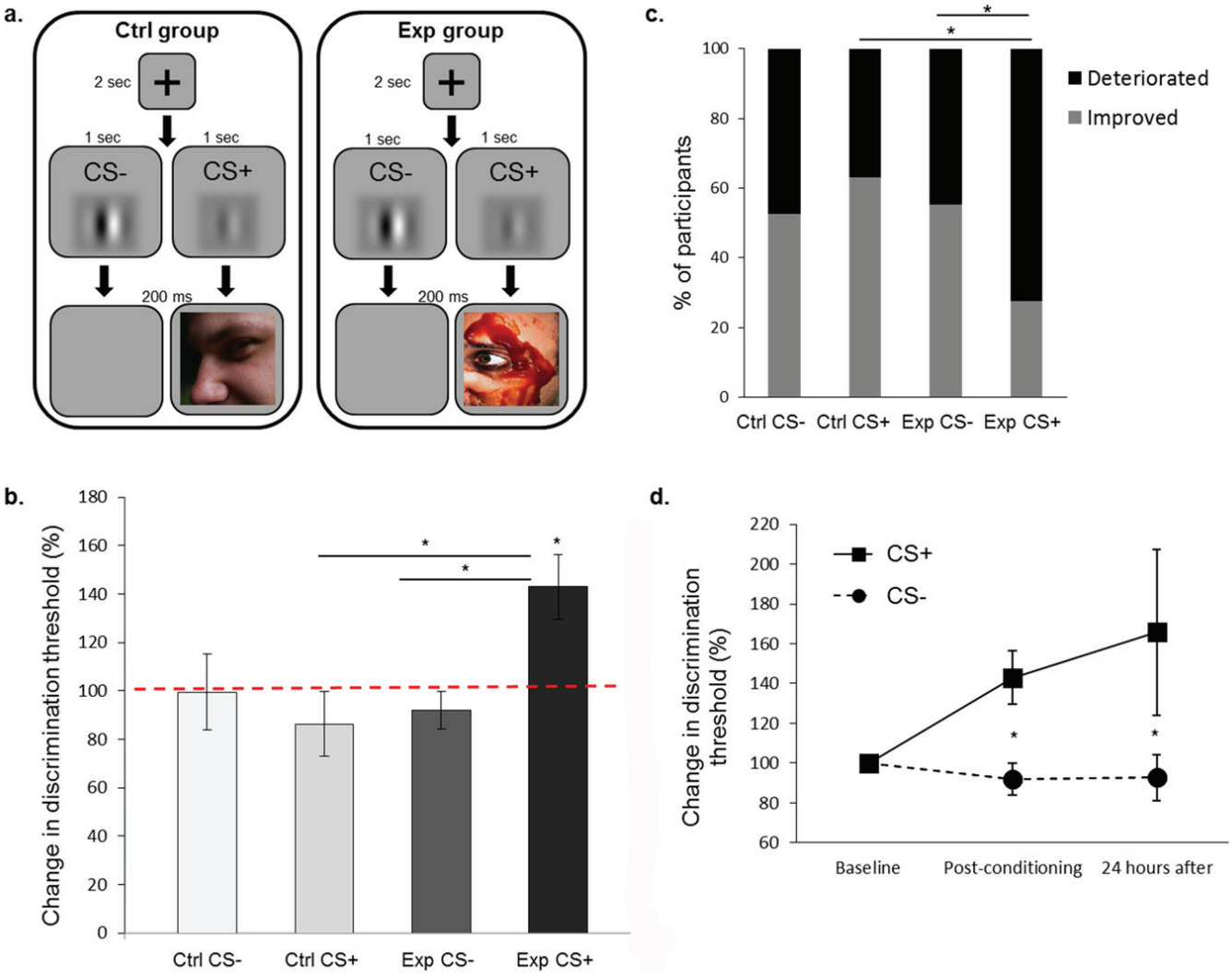




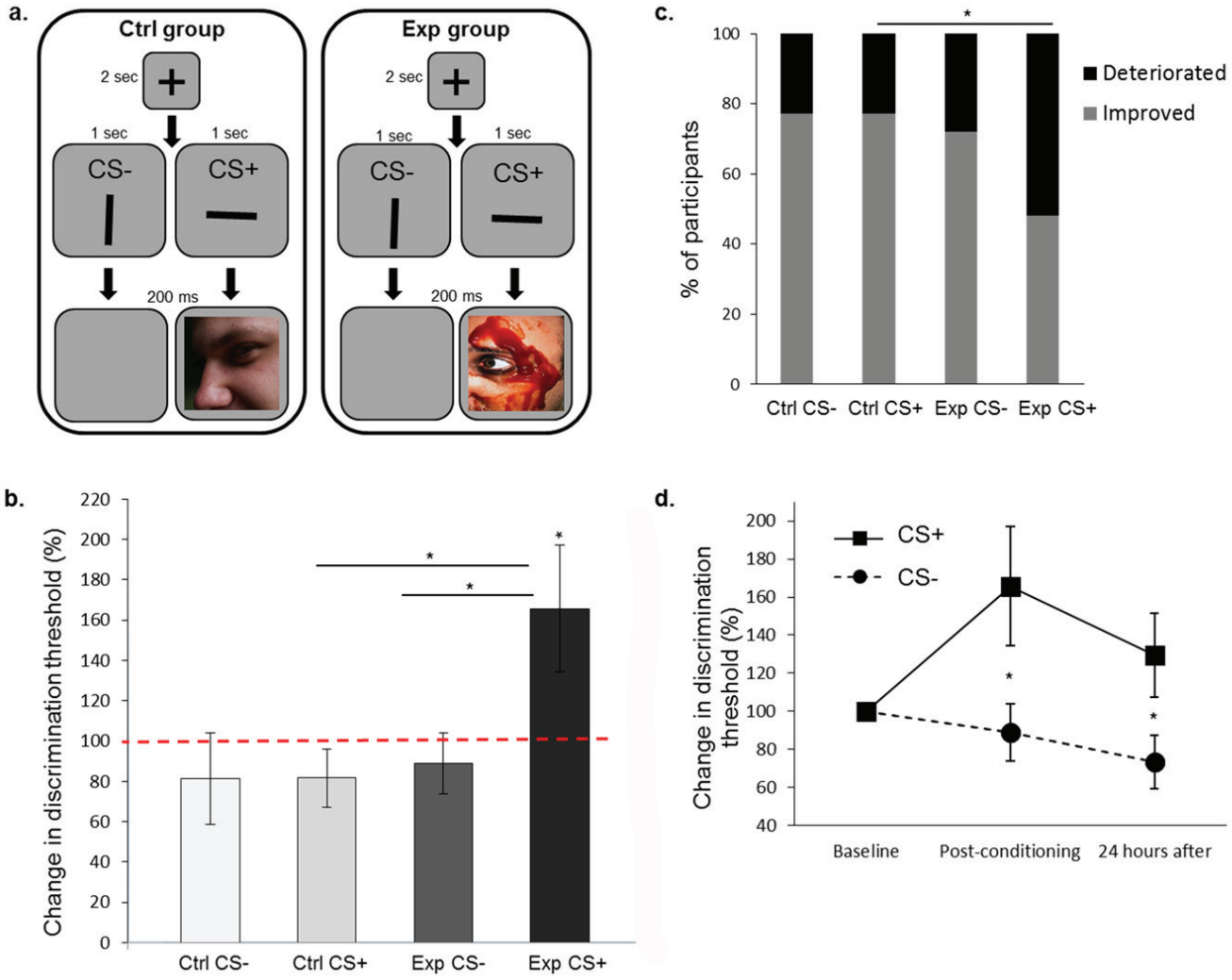




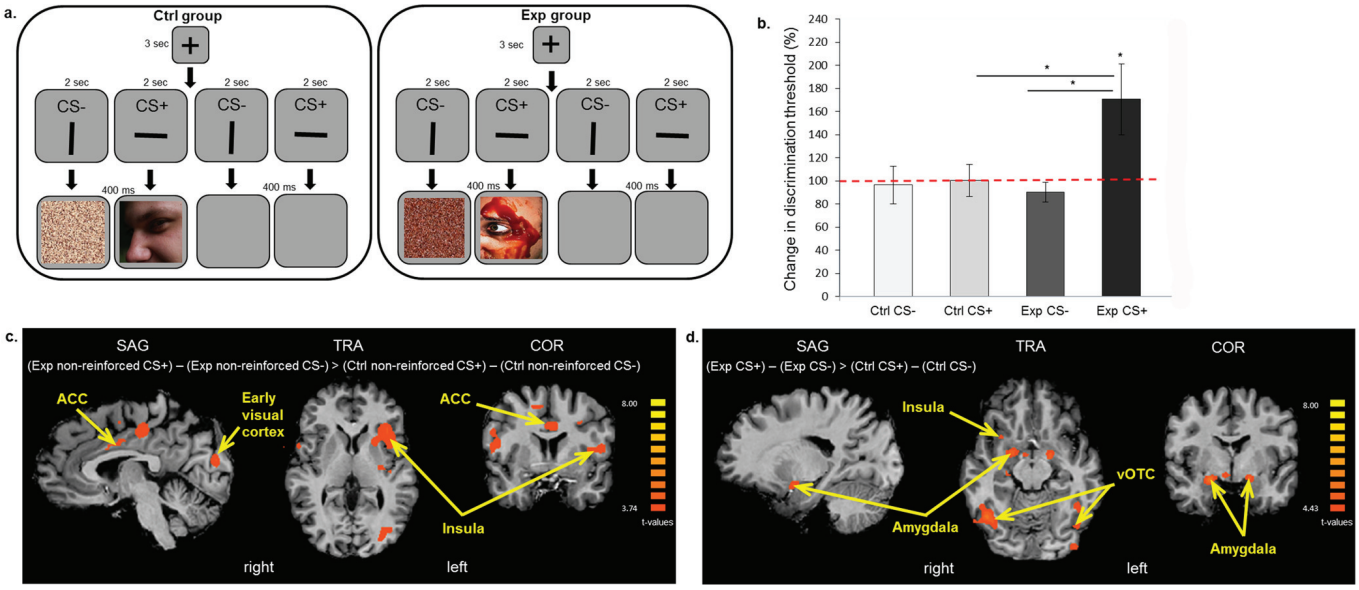
JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript