

Safety Signals in the Primate Amygdala

Rotem Genuad-Gabai, Oded Klavir, and Rony Paz

Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel 76100

The ability to distinguish danger from safety is crucial for survival. On the other hand, anxiety disorders can result from failures to dissociate safe cues from those that predict dangerous outcomes. The amygdala plays a major role in learning and signaling danger, and recently, evidence accumulates that it also acquires information to signal safety. Traditionally, safety is explored by paradigms that change the value of a previously dangerous cue, such as extinction or reversal; or by paradigms showing that a safe cue can inhibit responses to another danger-predicting cue, as in conditioned-inhibition. In real-life scenarios, many cues are never paired or tested with danger and remain neutral all along. A detailed study of neural responses to unpaired conditioned-stimulus (CS[−]) can therefore indicate whether information on safety-by-comparison is also acquired in the amygdala. We designed a multiple-CS study, with CS[−] from both visual and auditory modalities. Using discriminative aversive-conditioning, we find that responses in the primate amygdala develop for CS[−] of the same modality and of a different modality from that of the aversive CS⁺. Moreover, we find that responses are comparable in proportion, sign (increase/decrease), onset, and magnitude. These results indicate that the primate amygdala actively acquires signals about safety, and strengthen the hypothesis that failure in amygdala processing can result in failure to distinguish dangerous cues from safe ones and lead to maladaptive behaviors.

Introduction

Dissociating cues that predict danger from those that do not is necessary for normal behavior and survival. The amygdala was shown to have a crucial role in acquisition and maintenance of negative memories and emotions (Büchel and Dolan, 2000; LeDoux, 2000; Maren and Quirk, 2004; LaBar and Cabeza, 2006; Murray, 2007; Pape and Pare, 2010; Salzman and Fusi, 2010; Paz and Pare, 2013). On the other hand, failure to dissociate safe cues from dangerous ones produces inappropriate responses and might lead to anxiety and post-trauma disorders (PTSDs; van der Kolk, 1997; Kheirbek et al., 2012; Lissek, 2012; Pitman et al., 2012). Accordingly, aversive-conditioning results in less discrimination and wider generalization (Schechtman et al., 2010; Dunsmoor et al., 2011; Resnik et al., 2011; Laufer and Paz, 2012; Dunsmoor and LaBar, 2013), and PTSD and anxiety patients exhibit overgeneralized emotional responses to similar yet safe stimuli (Rauch et al., 2006; Lissek et al., 2010; Jovanovic et al., 2010, 2012; Levy-Gigi et al., 2012). Therefore, understanding how safety and danger are encoded in the amygdala is of primary importance.

The study of safety can rely on paradigms that change the acquired aversive-value of a cue, such as extinction or reversal (Schiller and Delgado, 2010; Milad and Quirk, 2012), showing that cortical-amygdala circuits underlie this learning (Phelps et al., 2004; Schiller et al., 2008; Livneh and Paz, 2012b; Milad and Quirk, 2012). Importantly, it was shown that the amygdala is actively involved in extinction (Herry et al., 2008, 2010; Amano et al., 2010; Pare and Duvarci, 2012), and that responses of

extinction-specific single neurons emerge during the learning (Herry et al., 2008; Livneh and Paz, 2012a). In another class of paradigms, safety is learned for a different stimulus, and this cue acts as a conditioned inhibitor for fear behavior (Christianson et al., 2012). In a recent study, Sangha et al. (2013) used a novel variation of the summation paradigm to show that amygdala neurons acquire such safety responses and there was some overlap between these neurons and those that displayed reward-related signals, as previously found in the amygdala (Salzman et al., 2007; Shabel and Janak, 2009). This demonstrates that safety signals in the amygdala are required for correct distinction (Rogan et al., 2005; Ostroff et al., 2010; Kazama et al., 2012).

In the current study, we asked whether responses to a neutral cue, a unpaired conditioned-stimulus (CS[−]) that was never paired with a reinforcer, develop as safety-by-comparison to the CS⁺, and develop even before this cue is tested as a safety cue; i.e., presented with an aversive reinforcer or cue. Although several studies quantified responses to the CS[−] (Collins and Paré, 2000), it was used mainly as control comparison to CS⁺. Here we provide a complete investigation across modalities (visual and auditory) and properties of the response (magnitude, direction, extent). Our results show that active responses to the safe cue develop in parallel and similar strength to those of the CS⁺, suggesting that the distinction between safety and danger is actively maintained in primate amygdala (BLA) circuits.

Materials and Methods

Animals. Two male *Macaca fascicularis* (4–7 kg) were implanted with a recording chamber (27 × 27 mm) above the right amygdala, under deep anesthesia and aseptic conditions. All surgical and experimental procedures were approved and conducted in accordance with the regulations of the Weizmann Institute Animal Care and Use Committee (IACUC), following NIH regulations and with AAALAC accreditation. Food, water, and enrichments (e.g., fruits and play instruments) were available *ad libitum* during the whole period, except before medical procedures.

Behavioral paradigm. During each experimental day the monkeys engaged in a 500 ms trace-classical-conditioning task with a random inter

Received April 10, 2013; revised Oct. 3, 2013; accepted Oct. 3, 2013.

Author contributions: R.G.-G., O.K., and R.P. designed research; R.G.-G. and O.K. performed research; R.G.-G. and R.P. analyzed data; R.G.-G. and R.P. wrote the paper.

The work was supported by ISF No. 26613 and ERC-FP7-StG No. 281171 grants to R.P. We thank Yossi Shohat for valuable contribution for the animals work and welfare, Dr Eilat Kahana and Dr Gil Hecht for help with medical and surgical procedures, and Dr Edna Furman-Haran and Nachum Stern for MRI procedures.

Correspondence should be addressed to Dr Rony Paz, Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel 76100. E-mail: rony.paz@weizmann.ac.il.

DOI:10.1523/JNEUROSCI.1539-13.2013

Copyright © 2013 the authors 0270-6474/13/3317986-09\$15.00/0

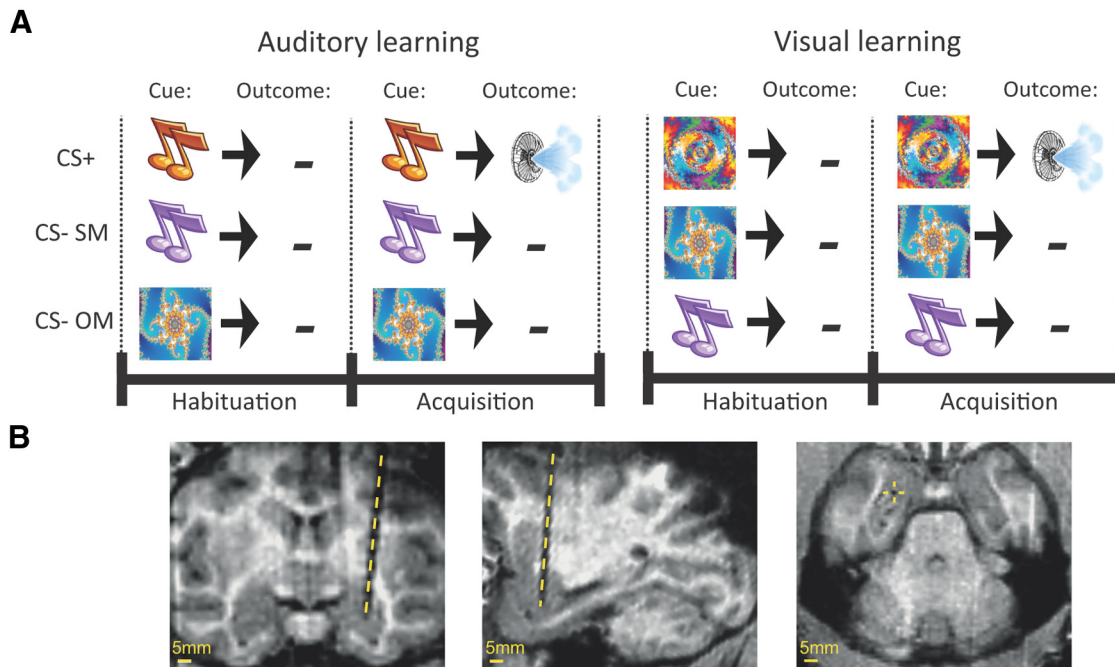


Figure 1. Experimental paradigm and recording site. **A**, Each session (day) included two phases: habituation, where stimuli were presented without any outcome (six trials per stimulus, pseudorandom); and Acquisition, where one stimulus (the CS+) was paired with an airpuff to the eye (CS+) and the two other stimuli (CS−) were unpaired (15 trials per stimulus). The CS+ was either an auditory pure tone (left columns) or a visual fractal (right columns), and one CS− was of the same modality (CS-SM) and the other of the other modality (CS-OM). Stimuli were presented for 2 s, followed by 0.5 s interval, and then reinforced (in CS+ trials), with 20–40 s ITI. In some sessions, a third phase was added (data not shown in the figure) in which the CS− became a CS+ to test that it acquired safety value. **B**, Anatomical MRI figures of the three planes (coronal, sagittal, horizontal) with an example electrode in the right amygdala (yellow). Recordings were verified and performed in the BLA with homogenous spread for lateral and basal nuclei (see text and Fig. 7).

trial interval (ITI) of 25 s on average (Fig. 1). The conditioned stimulus (CS+) with a duration of 2 s was either a new pure tone or a new fractal cue (Chaos Pro 4.0 program; www.chaospro.de), and was paired with an unconditioned aversive stimulus (US) of an air puff (150 ms duration; 3–5 bars; located proximally 5 cm from the left eye). Two other stimuli were presented without the aversive air puff outcome (CS−): one was a different pure tone and the other was a different fractal cue. Hence, in each day, there was an auditory or a visual CS+, and one CS− of the same sensory modality and one CS− of the other modality. Each new day started with a habituation phase where the three CSs were presented randomly (six trials each); followed by the acquisition phase described above (15 trials of each CS, pseudorandomly interleaved).

To verify that the CS− represents a safety signal after conditioning the CS+, an additional phase was added at the end in some sessions. In this phase (termed here “reversal” for clarity), the CS− was reinforced in a similar way to the CS+ during the previous conditioning phase. We then tested for rate of learning (“retardness”) and magnitude of responses at the end of this phase as evidence for safety (similar to conditioned- or latent-inhibition).

There were 100 experimental days overall in the two monkeys. Out of them, the CS+ was auditory in 53 d, and visual in 47 d. Auditory learning and visual learning days were randomly intermingled.

MRI-based electrode positioning. Anatomical MRI scans were acquired before, during, and after the recording period. Images were acquired on a 3-tesla MRI scanner: (MAGNETOM Trio, Siemens) with a CP knee coil (Siemens). T1 weighted and 3D gradient-echo (MPRAGE) pulse sequence was acquired with TR of 2500 ms, TI of 1100 ms, TE of 3.36 ms, 8° flip angle, and two averages. Images were acquired in the sagittal plane, 192 × 192 matrix and 0.8³ or 0.6³ mm resolution. A first scan was performed before surgery and used to align and refine anatomical maps for each individual animal (relative location of the amygdala and anatomical markers, such as the interaural line and the anterior commissure; confirmed using Martin and Bowden [2000] and Saleem and Logothetis [2007]). We used this scan to guide the positioning of the chamber on the skull at the surgery. After surgery we performed another scan with two electrodes directed toward the amygdala (Fig. 2), and 2–3 observers separately inspected the images and

calculated the amygdala anterior–posterior and lateral–medial borders relative to the electrodes. The depth of the amygdala was calculated from the dura surface.

Recordings. The monkeys were seated in a dark room and each day, 2–4 microelectrodes (0.6–1.2 MΩ glass/narylene-coated tungsten, Alpha Omega or We-sense) were lowered inside a metal guide (gauge 25xxtw, OD:0.51 mm, ID:0.41 mm; Cadence) into the brain using a head-tower and electrode-positioning-system (Alpha-Omega). The guide was lowered to penetrate and cross the dura and stopped once in the cortex. The electrodes were then moved independently further into the amygdala (we performed 4–7 mapping sessions in each animal by moving slowly and identifying electrophysiological markers of firing properties tracking the known anatomical pathway into the amygdala). Electrode signals were preamplified, 0.3–6 KHz bandpass filtered and sampled at 25 KHz, and on-line spike sorting was performed using a template-based algorithm (Alpha Lab Pro, Alpha Omega). We allowed 30 min for the tissue and signal to stabilize before starting acquisition and behavioral protocol. At the end of the recording period, off-line spike sorting was further performed for all sessions to improve unit isolation (offline sorter, Plexon).

Behavior. A computerized digital video camera for night conditions (Provision-Isr) recorded the monkey’s left eye at 50 Hz. Video analysis was performed on custom-made software implemented on Matlab to identify transitions from open to closed-state (an “eye-blink”). We validated the algorithm by random samples from several recording days and found it to be consistent with the judgments of a human observer for >95% of the reported eye-blinks. Quantifying behavior by the number of milliseconds in which the eye was closed did not change the main results. Number of eye-blinks was then normalized to baseline blink-rate in the period preceding CS presentation (−4 to −1 s).

Behavioral *d'* discrimination index was defined as:

$$d' = \frac{\text{blinks (CS+)} - \text{blinks (CS-)}}{\sqrt{\frac{\text{std}(\text{blinks (CS+)}) + \text{std}(\text{blinks (CS-)})}{2}}}$$

where blinks (CS) is the number of blinks during CS presentation and the first 400 ms of the trace-interval, taken from the last seven trials of the acquisition.

Statistical analysis. All results concerning behavior were analyzed using a repeated-measures ANOVA with eye-blink number as dependent variable. Significant interactions were followed by *post hoc* least significant difference comparisons, and for all comparisons, significance was assumed at $p < 0.05$. Days were defined as sessions with discriminative learning if the t test between CS+ and CS- reached significance.

For neural analysis, firing rates were calculated during CS presentation (2 s), and pre-CS (-4: -1 s before the CS), and normalized by the same periods averaged over the last three trials of the habituation phase. Only cells with overall average firing rate of >1 Hz were included in the analysis. To classify whether a neuron is CS- responsive, a Wilcoxon test compared pre-CS and post-CS from acquisition trials. Notice that the initial normalization to the end of habituation means that pure stimulus-evoked responses were not considered significant by themselves, and only changes due to learning were classified as such. Distribution statistics and proportions were analyzed using χ^2 test or Kolmogorov-Smirnov tests. Onsets were calculated using a custom-made implementation of the CUSUM approach, with two sigmas as confidence threshold.

Results

We tested amygdala neural activity during eye-blink trace-conditioning to a visual or an auditory CS. In each day the animals were first habituated to three new CSs; two of the same sensory modality and one of the other modality (six trials for each CS). During conditioning, one CS (CS+) was paired with an air puff in a 500 ms trace-conditioning paradigm, and the two other CSs used as CS- (Fig. 1A). Overall, there were 100 sessions, 24/23 visual in monkeys A/B and 29/24 auditory. Eye-blink aversive trace-conditioning was previously found to robustly recruit primate amygdala activity (Paton et al., 2006; Belova et al., 2007). We recorded single-unit activity (235 neurons, 125/110 in monkeys A/B) in verified recording locations in the BLA, confirmed by multiple MRI sessions (Fig. 1B).

Differential learning occurs for both modalities but with some within-modality generalization

We quantified learning by the number of eye-blinks normalized to baseline blink-rate in the period preceding CS presentation (-4 to -1 s). The probability for an eye blink increased during CS+ presentation with learning, reaching a maximum right after the CS ended but before the aversive US (the trace-interval), hence a preparatory response

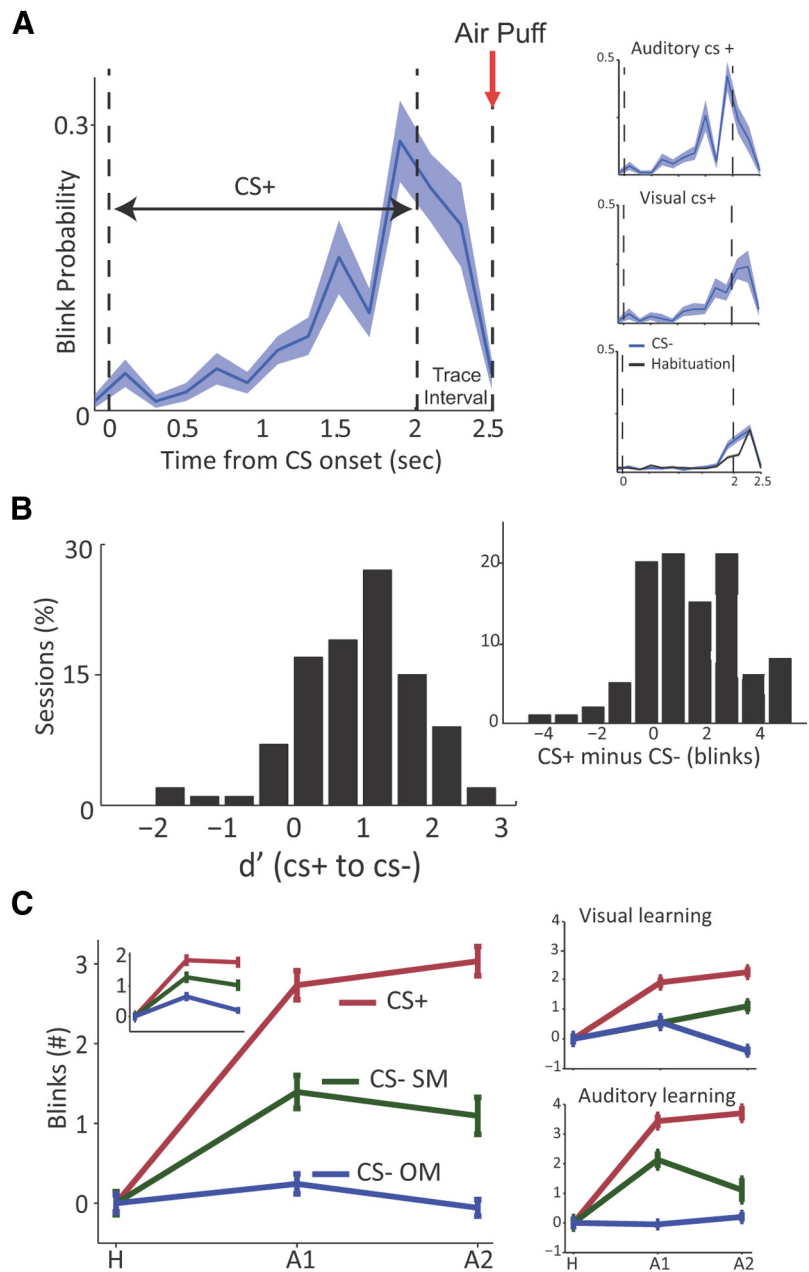


Figure 2. Learned behavior during discriminative aversive-conditioning. **A**, The CR revealed by the probability for closed-eye during CS+ trials at the end of learning (averaged over all sessions), starting 50 ms before stimulus onset, during the 2 s presentation, and in the 0.5 s between CS offset and airpuff delivery (red arrow at 2.5). A gradual increase can be observed during CS presentation and a marked additional increase during the interval preceding the airpuff; a preparatory response. Right, Top insets show the same data but separately for sessions of visual and auditory CS+. Right, bottom inset shows the response to the CS- and during habituation, showing natural blink response at CS termination. **B**, Distribution of behavioral d' from all sessions, calculated between the CS+ and CS- over the last seven trials of each session. The right inset shows the same data as difference in actual number of eye-blinks. Both distributions were significantly different from normal ($p < 0.01$, K-S test), and a more rigorous per-session test revealed significant differential learning in 21/47 visual days and 25/53 auditory days (t tests per day over the last 7 trials, $p < 0.05$). **C**, Mean number of CR (blinks \pm SEM) for the three types of CS (CS+, red; CS-SM, green; CS-OM, blue), at the end of habituation end (H; last five trials), Early acquisition (A1, first seven trials), and Late acquisition (A2, last seven trials). CRs are normalized to pre-CS period (-4 to -1 s) and to habituation, hence it represents the additional learned CR. Main figure includes all days that passed per-day differential learning test (**B**). Left, Inset includes all days; and two right insets separately per modality of CS+. Please see Results, Differential learning occurs for both modalities but with some within-modality generalization, for statistics: learning was robust for all conditions, with both across and within modality differential learning, and with some within-modality generalization when compared with habituation.

(Fig. 2A, right insets; for both visual and auditory CS). We then validated differential learning by calculating differences between blinks for the CS+ versus the CS- at the end of learning (last seven trials), and calculated behavioral d' scores for each

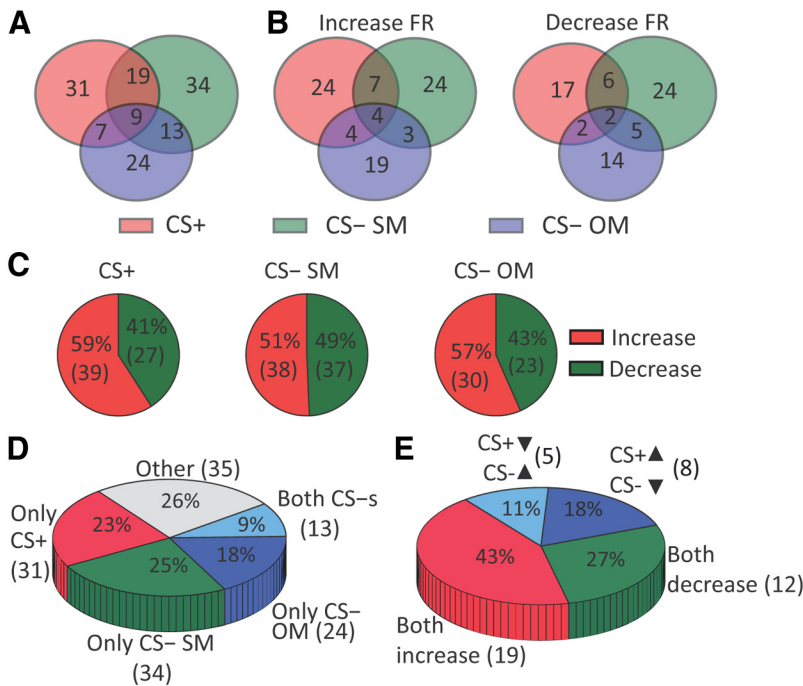


Figure 3. Distribution of acquired responses to the different CS types. A responsive neuron is tested for its firing rate following CS-onset but before the US-onset, compared with pre-CS activity, but importantly, after normalization to CS-evoked responses during habituation (see Materials and Methods). Hence, a responsive neuron is one that changed its activity following conditioning. **A**, Numbers of neurons per each CS (circle sizes are proportional): response to CS+ (red), CS– SM (green), CS– OM (blue). No significant difference was found (see Results, Proportion and sign of neural responses homogenously distribute across CS types). **B**, Same as in **A** but separately for neurons that increased (left) or decreased (right) their FR following CS onset. **C**, Dividing increases and decreases in FR for the neurons that acquired responses to each CS type. **D**, Distributions of exclusive acquired responses to each CS type. **E**, Directionality of responses divided by CS type: neurons that increase their FR to both CS+ and CS–, decrease to both, or in opposite directions. The take-home-message from these detailed careful categorizations is that there were little (nonsignificant) differences in proportions of acquired responses, and BLA neurons acquired increases and decreases in their FR to both CS+ and CS–. Hence, homogenous encoding is observed in the amygdala.

day/session, which were significantly shifted from normal (Fig. 2B; $p < 0.001$, Kolmogorov–Smirnov tests; right histogram shows the actual difference in average blinks per session, $p < 0.01$).

To quantify learning further and in detail, we normalized the blinks in the first seven trials (A1) and in the last seven trials (A2) of the acquisition to the habituation (H), and performed a complete three-way ANOVA on all experimental factors (Fig. 2C). Overall, there was a significant effect for phase (early/late acquisition, $F_{(2,411)} = 59.57, p < 0.01$), CS+ modality (visual/auditory, $F_{(1,412)} = 36.45, p < 0.01$), and CS category (CS– of the same modality or other modality, $F_{(2,411)} = 23.28, p < 0.01$). There was also a significant interaction between CS+ modality and CS category ($F_{(2,408)} = 53.51, p < 0.01$), between CS+ modality and phase ($F_{(2,408)} = 3.85, p < 0.05$) and between CS category and phase ($F_{(4,405)} = 8.1, p < 0.01$).

Importantly, *post hoc* least significant difference comparisons (PLSD) revealed a significant higher response for CS+ compared with both CS– ($p < 0.05$; Fig. 2C), indicating that learning occurred robustly for the CS+. However, we observed within-modality generalization, as response for CS+ and CS– of the same modality were both significantly higher compared with habituation ($p < 0.05$); but response for the CS– of the other modality was not ($p > 0.05$; Fig. 2C). In other words, whereas the CS+ was differentiated from both CS–, only the CS– of the other modality returned to habituation levels, whereas that of the same modality still evoked preparatory conditioned responses (CRs).

Proportion and sign of neural responses homogenously distribute across CS types

Of 235 recorded neurons (125/110 from monkeys A/B, respectively), 137 acquired responses to one or more of the CSs. All reported responses from here on refer to learned responses, i.e., responses that were acquired during learning; because tests were performed after normalization to habituation and by comparing pre-CS baseline to CS-elicited response ($p < 0.05$, Wilcoxon; see Materials and Methods). Overall, 66 neurons acquired responses to the CS+, 75 neurons acquired responses to CS– of the same modality, and 53 to CS– of the other modality. Perhaps surprisingly, these proportions were homogenously distributed across the three different CS categories (Fig. 3A; $p > 0.05, \chi^2$).

We further divided responses to increasing and decreasing firing rates (FRs; Fig. 4A), compared with baseline and habituation. Of the 107 neurons that increased their FR, 39 did so for the CS+, 38 to the CS– of the same modality, and 30 to the CS– of the other modality (Fig. 3B, left). Of the 87 neurons that decreased FR, 27 and 37 did so for the CS+ and CS– of the same modality respectively, and 23 for the CS– of the other modality (Fig. 3B, right). As can be seen, here again, responses were homogenously distributed across the 3 CSs ($p > 0.05$ for both increased and decreased FR, χ^2). Moreover, learned responses to the different CSs were homogenously distributed between increasing and decreasing FRs (Fig. 3C; $p > 0.05$ for all, χ^2).

To make an even finer analysis, we tested the distribution of specific responses: neurons that acquired responses to only one CS; or to both CS– but not CS+ (Figs. 3D, 4B). Neurons that respond to the CS+ alone comprised 23% of all responsive neurons; 25% responded only to CS– of the same modality, and 18% to CS– of the other modality. These exclusive responses were again homogenously distributed ($p > 0.05, \chi^2$). Nine percent (9%) of the neurons responded to both CS– but not to the CS+, and the rest (26%) responded to all CSs or to CS+ and one of the CS– (Fig. 3D).

Finally, we tested whether neurons that acquired responses to both the CS+ and a CS– (44 neurons), did so in similar directions, i.e., increase/decrease FR. We found that most neurons (70.5%) acquired responses to CS+ and CS– in the same direction: 43% increased FRs for both and 27% decreased for both CS+ and CS–. Responses in opposite directions for CS+ and CS– were observed in 29.5% of the cases; 18% increased FRs for the CS+ and decreased for the CS–, and 11% vice versa (Fig. 3E).

To conclude, it seems that single-cells in the primate amygdala acquire responses in similar proportions to CS+ and CS– during discriminative learning; and moreover, they do so in similar proportions of increases and decreases of their FR.

Characteristics of neural responses to the different CSs

We tested response characteristics, to further identify differences in acquired responses for the different CS categories.

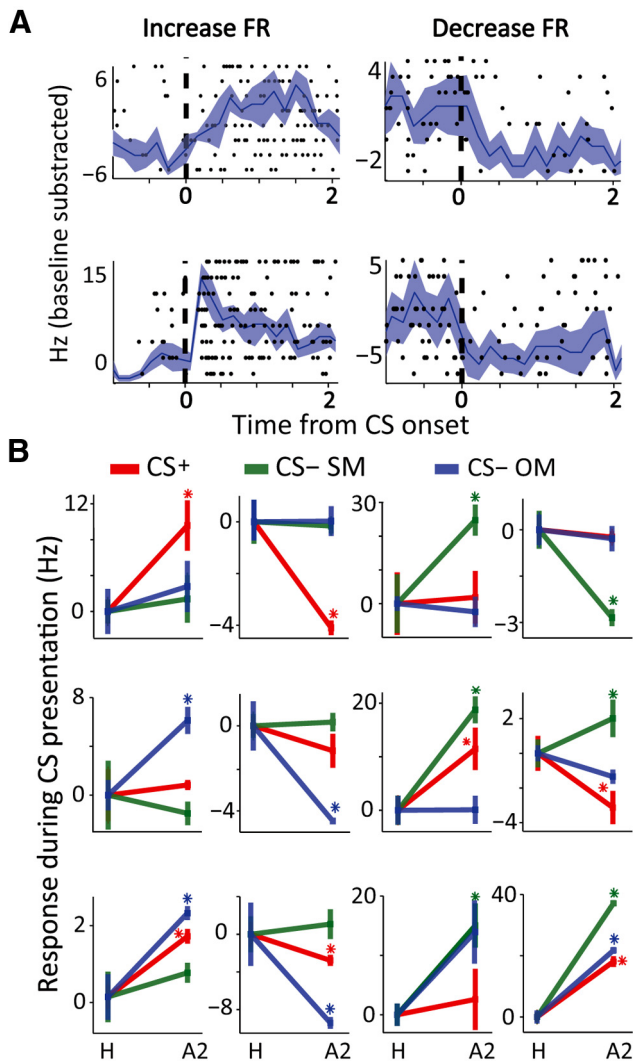


Figure 4. Neuronal responses. **A**, Raster plots overlaid with PSTHs for four neurons, with tonic increase in FR (top left), phasic increase (bottom left), or decrease (right) their FR following CS presentation (dashed line indicates CS onset). **B**, Twelve different examples of individual neurons and their acquired response (spikes/s, mean \pm SEM), comparing habituation (H) to last seven trials of acquisition (A). Firing rate was normalized to pre-CS baseline (-4 to -1 s) and to habituation, hence this is learned-acquired response. There is large heterogeneity in responses for CS types. The proportions are explained and summarized in Fig. 3.

We first compared the magnitude of FR acquired response. There was no significant difference in the mean FR between neurons that responded to different CSs, for both decreases and increases (Fig. 5A,B; $p > 0.05$, ANOVA). This was true both for CS-evoked responses at the end of learning that were not normalized and baseline subtracted (data not shown; $p > 0.05$, ANOVA), and for normalized acquired responses (Fig. 5B; $p > 0.05$, ANOVA). Similar results were found when analyzing separately visual learning days and auditory learning days (data presented per modality; Fig. 5).

We then compared onsets of neural activity to the different CSs at habituation and at the end of acquisition, for both modalities of CS+. A three-way ANOVA found a significant effect of learning phase (Fig. 6B; $F_{(1,689)} = 5.851$, $p = 0.02$), no effect for CS+ modality (Fig. 6A; $F_{(1,689)} = 0.06$, $p = 0.8$), and no effect of CS- category ($F_{(2,688)} = 2.619$, $p = 0.07$; Fig. 6C). There were no interactions ($F_{(2,685)} = 0.28$, $p = 0.7$; $F_{(2,685)} = 1.02$, $p = 0.4$; $F_{(1,687)} = 0.01$, $p = 0.9$). We calculated the difference between

onsets at the end of learning to those during habituation for the different CSs, and plotted distributions of all responsive neurons (Fig. 6D; $F_{(2,148)} = 2.054$, $p > 0.1$). Even a more sensitive model-free test (two-sample Kolmogorov–Smirnov) found no difference between the distributions of onsets for the three CS types (Fig. 6E; $p > 0.05$, CS+ vs CS- of the same modality, $p > 0.2$; CS+ vs CS- of the other modality, $p > 0.3$; CS- of the same modality vs CS- of the other modality, $p > 0.3$).

Finally, we tested whether there is a difference in the anatomical position within the BLA of responsive neurons to the different CSs. To do this, we reconstructed all recording locations and divided responsive cells along the three possible dimensions: ventral–dorsal, medial–lateral, and anterior–posterior. Neurons that acquired responses to all three CS types (CS+, CS- of the same modality and CS- of the other modality) were homogeneously distributed anatomically throughout the BLA (Fig. 7; χ^2 tests, $p > 0.05$ for all).

Reduced safety representation underlies generalized behavior

We defined sessions with discriminative learning if the t test between CS+ and CS- reached significance at the end of learning (last seven trials, t test comparing eye-blinks, $p < 0.05$). We found that during discriminating days ($n = 46$), there were similar numbers of neurons responding to the CS+ and to the CS- ($n = 20$ and $n = 22$, $p > 0.1$, χ^2), suggesting that indeed safety-coding in the BLA is required for successful discrimination. In contrast, during sessions in which discrimination was not significant, there were more neurons responding to the CS+ than to the CS- ($n = 27$, $n = 15$, $p < 0.05$, χ^2). In addition, comparing FRs showed that these were increased during nondiscriminating days (5.6 ± 1.0 for CS+ and 4.8 ± 1.0 for CS-) versus discriminating days (3.0 ± 0.6 for CS+ and 3.2 ± 0.6 for CS-), but there were no differences between the CS+ and the CS- within the discriminating or the nondiscriminating days (two-way ANOVA revealed only a main effect of discriminating versus nondiscriminating sessions, no main effect of CS type, and no interaction effect). These results could further suggest that CS- responding neurons encode safety, and in days where less neurons responded to the CS- and FR was overall elevated, behavior showed generalized anxiety to both CS.

Discriminative conditioning induce safety learning

We verified that the CS- acquired a safety value when the CS+ was conditioned (“safety-by-comparison”). To test this, we performed sessions where at the end of the conditioning phase, an additional phase occurred in which the CS- was reinforced (termed here reversal for clarity). The CS- (now CS+) was conditioned for a similar number of trials (15) as in the previous acquisition session. We then tested for rate of learning (retardness) and magnitude of responses at the end of the session, similar to probing conditioned-inhibition (Rescorla, 1969; Christianson et al., 2012) or latent-inhibition (Lubow and Moore, 1959). We found that learning was slower for the CS-, when compared with the CS+ from the previous acquisition phase, and responses were reduced at its end (Fig. 8A; two-way ANOVA with significant effect for both acquisition/reversal and early/late, and significant interaction, all $p < 0.01$; PLSD showed significant difference in both early and late phases when comparing acquisition to reversal, $p < 0.05$). This suggests that the CS- induced safety learning, when combined in a discriminative aversive learning paradigm. Moreover, analyzing neurons that were recorded in these sessions showed that of 57 neurons that responded to the CS- at the end of acquisition, 72% (41, $p <$

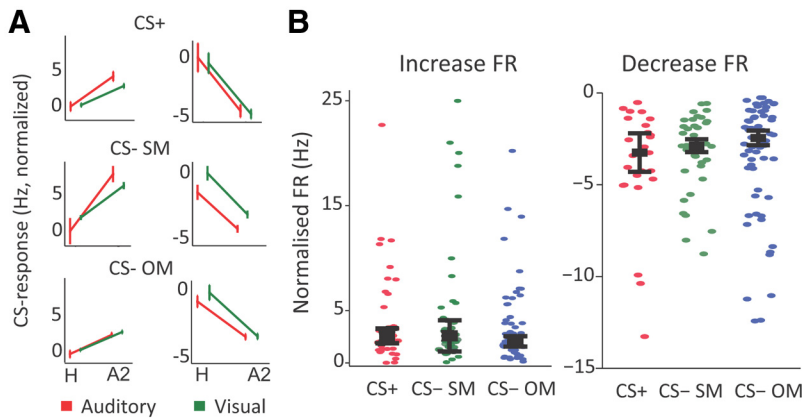


Figure 5. Magnitudes of responses. **A**, Acquired responses, averaged over all neurons (mean ± SEM, after normalization) for the different CS types (rows), separately for increasing and decreasing neurons (left and right columns), and for visual or auditory sessions (green and red). See Results, Characteristics of neural responses to the different CSs, for statistics. **B**, FR from all responsive neurons, separately for increasing and decreasing FR (left and right plots) and the three CS types (different colors and columns within each plot). There were no significant differences in the magnitude of the population response to the different CS types.

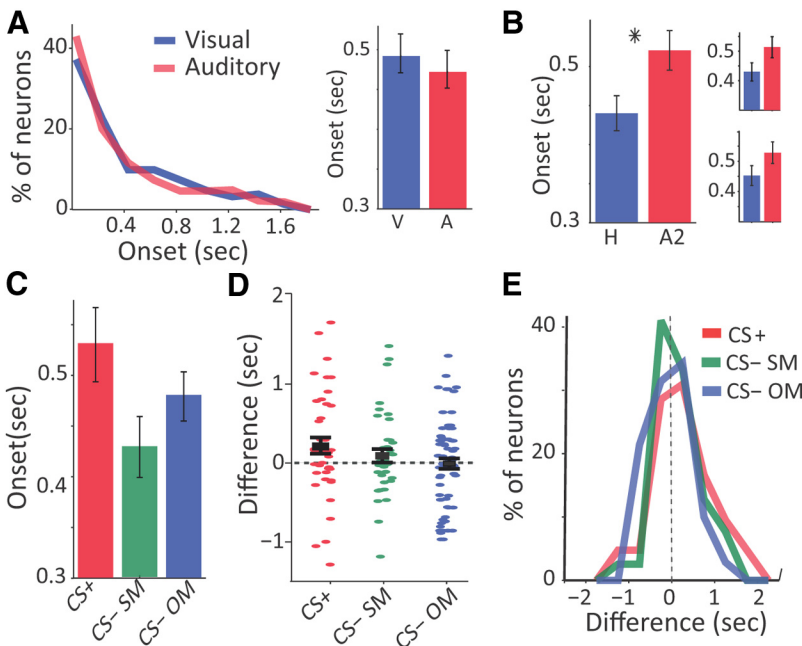


Figure 6. Onset of responses. **A**, Distribution of onsets for neural responses from all responsive neurons, separately for auditory and visual sessions. Mean ± SEM shown to the right. **B**, Onsets mean ± SEM during habituation (H) and late acquisition (A2). Insets in the right show the same separately for auditory and visual sessions. **C**, Onsets mean ± SEM for the different CS types (averaged over H and A2). **D**, The difference between onsets of responding neurons at acquisition and habituation for the different CS types. Shown in black is mean ± SEM. **E**, Distribution of onset differences (same data as in **D**). Whereas onsets were delayed after learning (**C**), potentially reflecting the learned delay (trace interval and/or CS duration), there were little differences between CS types and modalities (see Results, Characteristics of neural responses to the different CSs, for statistics).

0.01, χ^2) decreased their response (26 stopped responding to the CS completely, and 15 decreased their FR) when the CS– turned into a CS+ during the reversal (Fig. 8B). This further suggests that CS– exclusive neurons indeed signal safety during acquisition.

Discussion

This study was designed to explore in detail acquired and learned neural responses in the amygdala to an unpaired stimulus; i.e., a safe cue, one that signals the absence of any outcome. We recorded 235 neurons in the BLA of two behaving nonhuman primates

during complex discriminative aversive-conditioning, that includes two modalities as the CS+ (auditory or visual, in different sessions), and with two different CS–, one of the same modality as of the CS+ and one of the other modality. We then performed a rigorous detailed analysis of acquired responses and their characteristics, so that we can directly compare, in detail and large-scale statistics, the changes for the aversive conditioned stimulus versus those for the two safe ones. We found that the primate BLA robustly codes and acquires responses to the safety predicting stimuli as well. Perhaps surprisingly, we found similar extents of changes in the neural properties, such as proportions of responsive neurons, magnitude of responses, increases and decreases, onsets, and anatomical position. These results have direct implications to the delicate balance between negative (e.g., fear) and safety responses driven by amygdala circuitry. Below, we discuss several aspects of these findings.

A recent study demonstrated safety encoding in the rat amygdala (Sangha et al., 2013). By using an elegant adaptation of the summation test, they showed that neurons in the basal amygdala respond to safety cues (a CS– presented in combination with a CS+), and moreover, they found that some of these neurons also respond to reward cues, whereas others do not (suggesting further that avoiding negative valence is rewarding; Kim et al., 2006). Our study confirms this robust encoding of safety signals in the amygdala, and adds to it in several aspects: first, we found this in the primate BLA, extending the translational value and implications to anxiety-disorders (further discussed later); second, we recorded during acquisition, and performed all analyses on this phase. This approach provides strong confirmation that safety encoding develops independently of the behavioral test for safety. In the summation test, for example, the CS– and CS+ are presented in parallel, and therefore the neural findings could be affected by the fact that the animal was exposed to potential association between the two (which could result in an aversive value assigned to the CS–, or reducing the aversiveness of the CS+, as indeed is the purpose of this test). Although we do not think this is a confound for the interpretation of safety encoding as in Sangha et al. (2013), our study extends and confirms that robust encoding develops to the CS– even before the animal experienced explicit safety testing.

Third, we expanded here on the range of characteristics of neural activity analyzed, and showed that not only there is robust coding to a safety cue, but it is also extensive and comparable in magnitude, onsets, and directionality. This could seem at first to be in some discrepancy with a previous study that reported in-

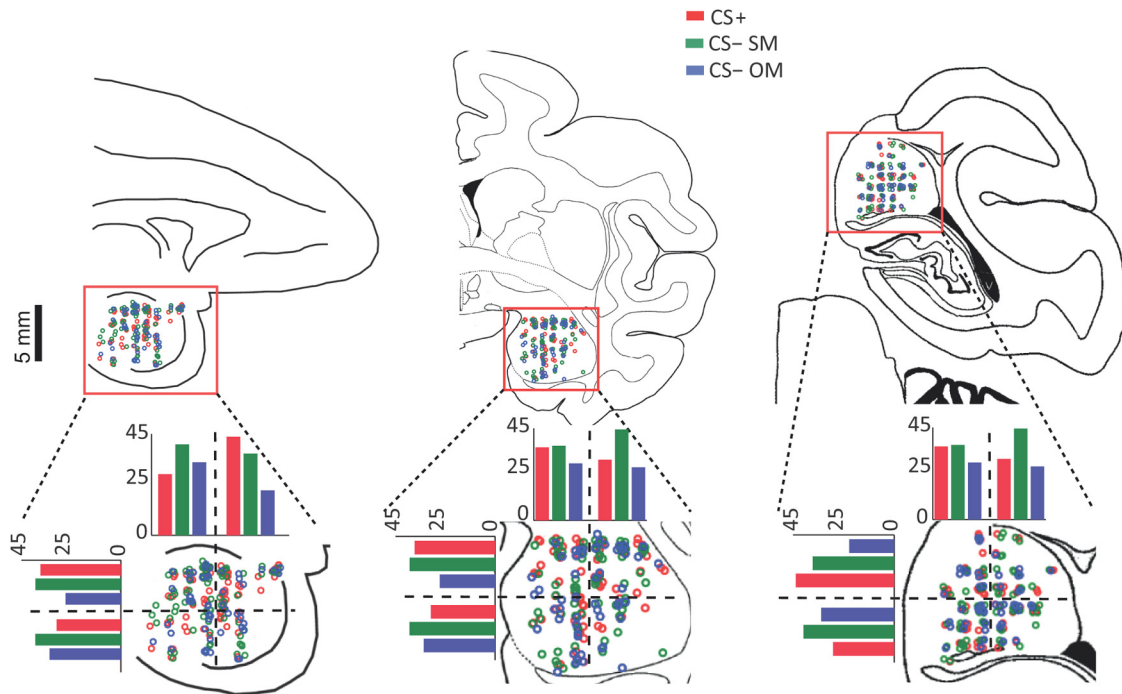


Figure 7. Recording sites and anatomical distribution of responding neurons. Schemes of anatomical sections of the macaque brain (sagittal, coronal and horizontal), with recording sites reconstructed based on multiple MRI with calibrating electrodes. Distribution of responsive neurons to the different CS types is shown at each section (CS+, red; CS– same modality, green; CS– other modality, blue). Bar graphs show percentage of responsive neurons in each half section.

creases to the CS+ versus decreases to the CS– (Collins and Paré, 2000). Although differences could be attributed to experimental paradigm, e.g., their use of the more intensive electrical shock or to species differences and the wealth of cortical and sensory inputs to the BLA of primates versus cats; we believe a simpler more plausible explanation is the focus of analysis. Whereas they focused on the subpopulation that responded to the CS+ and tested how it responds on pooled average to the CS–, we performed complete analysis of all neurons, separating them into specific groups and identifying all subpopulations that respond to different cue combinations, and then examined their properties (directionality, onset, magnitude). The finding of separate subpopulations that respond to negative (aversive) or positive (safety) predicting cue is in-line with several recent studies, as detailed below.

We used a “retardation” test to show that safety was formed for the CS–. This is a classical finding (Rescorla, 1969), but one with meaning: many (basically all) surrounding stimuli are unpaired with the reinforcer, and hence the animal could have learned many things. The fact that we observed robust and widespread responses that develop to the two CS–, means that safety was learned by comparison, and some understanding occurred that these other two stimuli are more than “merely” neutral, but are real safety. One possible mechanism is that of active inhibition, as assumed for explicit safety (Christianson et al., 2012). Evidence pointing in this direction comes from the behavioral responses in our paradigm, that first occurred (increased) for all stimuli, and complete differentiation formed gradually during acquisition. The circuitry that controls these differences, i.e., what stimuli continue to be treated as neutral, and which actively becomes safety, is still unclear. A natural explanation is that of similarity, if a stimulus is “close” enough, the animal must learn to actively distinguish it. Such similarity can be caused by contextual cues, as in different/similar environments, structural similarities,

e.g., the stimulus is presented in the same form of trials (i.e., CS–); or it can be perceptual similarities, as with the two CS– of the same modality here. Much evidence implicates the hippocampus and prefrontal regions in contextual generalization (Eichenbaum, 2000; Maren and Quirk, 2004; Wiltgen and Silva, 2007; Wimmer and Shohamy, 2012; Xu and Südhof, 2013), and the amygdala with generalization of aversive learning (Armony et al., 1997; Laxmi et al., 2003; Shaban et al., 2006; Antunes and Moita, 2010; Laufer and Paz, 2012). Because we found responses to all cues, both CR-evoking and non-evoking, these findings suggest that amygdala dysfunction can explain differences in generalization to simple stimuli. It remains to be explored what conditions alter the balance in this emotional triad of hippocampus–amygdala–PFC, to generate generalization in different circumstances.

Here, we observed some within-modality generalization: although behavior was differentiated for the CS+ and CS– of the same modality at the end of learning, the CS– of the other modality usually returned to baseline (habituation) levels, whereas the CS– of the same modality still evoked responses. On one hand, this allowed us an internal control, because we could verify that neural responses develop not only to one CS– that is differentiated from the CS+ but still evokes some CR (as in most studies), but also to a CS– that becomes completely safe. This therefore further strengthens the result that although no CR is elicited, the amygdala still signals and encodes this cue. On the other hand, the generalized responses to the CS– of the same modality is in-line with recent studies that show that generalization of aversive learning can occur at perceptual levels (Resnik et al., 2011; Laufer and Paz, 2012; Dunsmoor and LaBar, 2013), and point to the amygdala as a source, probably within a circuit with prefrontal and even sensory cortices involved.

The fact that safety signals reside in the amygdala, and that we found a large subpopulation of neurons responding to CS– exclusively, is in-line with the idea that safety can be represented in

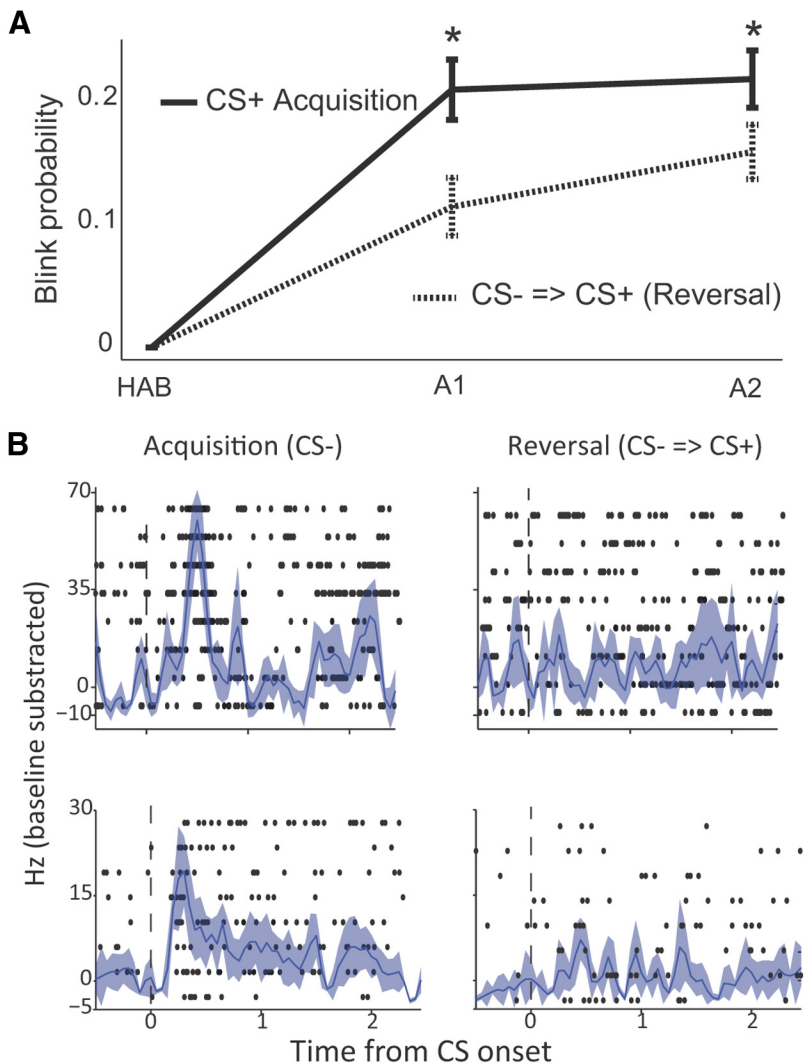


Figure 8. Delayed acquisition of CS⁻ implies safety learning. **A**, An additional phase was performed where the CS⁻ was now reinforced. Shown is a direct comparison between acquired CRs (blinks) during the acquisition phase for the CS⁺ (solid line), and during the reversal phase for the CS⁻, dashed line. Learning was delayed and reduced, implying the CS⁻ acquired safety value. **B**, Raster plots overlaid with PSTHs for two different neurons (rows) that signal safety during late acquisition (left column, response to the CS⁻), but stop responding when this cue becomes a CS⁺ during reversal (right column).

overlap to positive signals. This was demonstrated explicitly in (Sangha et al., 2013) study, confirming previous suggestions (Rogan et al., 2005; Shabel and Janak, 2009). The large population of cells responding to the two CS⁻ in our study, if interpreted as a positive thing for the animal (safety), is also in-line with previous studies showing widespread (in size and anatomical location) responses to cues predicting reward on top of the classical role of the amygdala (Baxter and Murray, 2002), and in-line with the fact that positive and negative signals are represented in subpopulations within the amygdala (Paton et al., 2006; Belova et al., 2007). Further, extinction, a form of safety (Schiller and Delgado, 2010; Milad and Quirk, 2012), is represented in distinct but widespread subpopulations of amygdala neurons (Herry et al., 2008; Livneh and Paz, 2012a). The overlap of different forms of safety and reward circuitry remains to be thoroughly examined (Schiller et al., 2008; Schiller and Delgado, 2010; Tye et al., 2010; Li et al., 2011; Christianson et al., 2012).

In sum, we conclude that the primate amygdala signals not only cues that predict danger, but also cues that predict safety. These responses could be driven by prefrontal neurons and regions that

were shown to be involved in emotion modulation and control (Ochsner and Gross, 2005; Quirk and Beer, 2006; Schiller and Delgado, 2010; Christianson et al., 2012; Klavir et al., 2012; Courtin et al., 2013), but could also be learned and stored locally overtime. Therefore, disruption in amygdala circuitry or prefrontal–amygdala connectivity could well lead to confusion and maladaptive responses, such as in the case of anxiety-disorders in general. Our findings suggest that even local dysfunction can result in inappropriate responses and generalization of anxiety and fear (van der Kolk, 1997; Dunsmoor et al., 2009; Lissek et al., 2010; Resnik et al., 2011; Jovanovic et al., 2012; Kheirbek et al., 2012; Laufer and Paz, 2012; Lissek, 2012; Pitman et al., 2012). Future studies will need to unveil the network architecture that can lead to normal versus abnormal responses.

References

- Amano T, Unal CT, Paré D (2010) Synaptic correlates of fear extinction in the amygdala. *Nat Neurosci* 13:489–494. [CrossRef Medline](#)
- Antunes R, Moita MA (2010) Discriminative auditory fear learning requires both tuned and nontuned auditory pathways to the amygdala. *J Neurosci* 30:9782–9787. [CrossRef Medline](#)
- Armony JL, Servan-Schreiber D, Romanski LM, Cohen JD, LeDoux JE (1997) Stimulus generalization of fear responses: effects of auditory cortex lesions in a computational model and in rats. *Cereb Cortex* 7:157–165. [CrossRef Medline](#)
- Baxter MG, Murray EA (2002) The amygdala and reward. *Nat Rev Neurosci* 3:563–573. [CrossRef Medline](#)
- Belova MA, Paton JJ, Morrison SE, Salzman CD (2007) Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55:970–984. [CrossRef Medline](#)
- Büchel C, Dolan RJ (2000) Classical fear conditioning in functional neuroimaging. *Curr Opin Neurobiol* 10:219–223. [CrossRef Medline](#)

Medline

- Christianson JP, Fernando AB, Kazama AM, Jovanovic T, Ostroff LE, Sangha S (2012) Inhibition of fear by learned safety signals: a mini-symposium review. *J Neurosci* 32:14118–14124. [CrossRef Medline](#)
- Collins DR, Paré D (2000) Differential fear conditioning induces reciprocal changes in the sensory responses of lateral amygdala neurons to the CS(+) and CS(-). *Learn Mem* 7:97–103. [CrossRef Medline](#)
- Courtin J, Bienvenu TC, Einarsson EÖ, Herry C (2013) Medial prefrontal cortex neuronal circuits in fear behavior. *Neuroscience* 240:219–242. [CrossRef Medline](#)
- Dunsmoor JE, LaBar KS (2013) Effects of discrimination training on fear generalization gradients and perceptual classification in humans. *Behav Neurosci* 127:350–356. [CrossRef Medline](#)
- Dunsmoor JE, Mitroff SR, LaBar KS (2009) Generalization of conditioned fear along a dimension of increasing fear intensity. *Learn Mem* 16:460–469. [CrossRef Medline](#)
- Dunsmoor JE, Prince SE, Murty VP, Kragel PA, LaBar KS (2011) Neurobehavioral mechanisms of human fear generalization. *Neuroimage* 55:1878–1888. [CrossRef Medline](#)
- Eichenbaum H (2000) A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci* 1:41–50. [CrossRef Medline](#)
- Herry C, Ciocchi S, Senn V, Demmou L, Müller C, Lüthi A (2008) Switching

- on and off fear by distinct neuronal circuits. *Nature* 454:600–606. [CrossRef Medline](#)
- Herry C, Ferraguti F, Singewald N, Letzkus JJ, Ehrlich I, Lüthi A (2010) Neuronal circuits of fear extinction. *Eur J Neurosci* 31:599–612. [CrossRef Medline](#)
- Jovanovic T, Norrholm SD, Blanding NQ, Davis M, Duncan E, Bradley B, Ressler KJ (2010) Impaired fear inhibition is a biomarker of PTSD but not depression. *Depress Anxiety* 27:244–251. [CrossRef Medline](#)
- Jovanovic T, Kazama A, Bachevalier J, Davis M (2012) Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology* 62:695–704. [CrossRef Medline](#)
- Kazama AM, Heuer E, Davis M, Bachevalier J (2012) Effects of neonatal amygdala lesions on fear learning, conditioned inhibition, and extinction in adult macaques. *Behav Neurosci* 126:392–403. [CrossRef Medline](#)
- Kheirbek MA, Klemenhagen KC, Sahay A, Hen R (2012) Neurogenesis and generalization: a new approach to stratify and treat anxiety disorders. *Nat Neurosci* 15:1613–1620. [CrossRef Medline](#)
- Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 4:e233. [CrossRef Medline](#)
- Klaviv O, Genud-Gabai R, Paz R (2012) Low-frequency stimulation depresses the primate anterior-cingulate-cortex and prevents spontaneous recovery of aversive memories. *J Neurosci* 32:8589–8597. [CrossRef Medline](#)
- LaBar KS, Cabeza R (2006) Cognitive neuroscience of emotional memory. *Nat Rev Neurosci* 7:54–64. [CrossRef Medline](#)
- Laufer O, Paz R (2012) Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. *J Neurosci* 32:6304–6311. [CrossRef Medline](#)
- Laxmi TR, Stork O, Pape HC (2003) Generalisation of conditioned fear and its behavioural expression in mice. *Behav Brain Res* 145:89–98. [CrossRef Medline](#)
- LeDoux JE (2000) Emotion circuits in the brain. *Annu Rev Neurosci* 23:155–184. [CrossRef Medline](#)
- Levy-Gigi E, Kéri S, Myers CE, Lencovsky Z, Sharvit-Benbaji H, Orr SP, Gilbertson MW, Servatius RJ, Tsao JW, Gluck MA (2012) Individuals with posttraumatic stress disorder show a selective deficit in generalization of associative learning. *Neuropsychology* 26:758–767. [CrossRef Medline](#)
- Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci* 14:1250–1252. [CrossRef Medline](#)
- Lissek S (2012) Toward an account of clinical anxiety predicated on basic, neurally mapped mechanisms of Pavlovian fear-learning: the case for conditioned overgeneralization. *Depress Anxiety* 29:257–263. [CrossRef Medline](#)
- Lissek S, Rabin S, Heller RE, Lukenbaugh D, Geraci M, Pine DS, Grillon C (2010) Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *Am J Psychiatry* 167:47–55. [CrossRef Medline](#)
- Livneh U, Paz R (2012a) Aversive-bias and stage-selectivity in neurons of the primate amygdala during acquisition, extinction, and overnight retention. *J Neurosci* 32:8598–8610. [CrossRef Medline](#)
- Livneh U, Paz R (2012b) Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. *Neuron* 75:133–142. [CrossRef Medline](#)
- Lubow RE, Moore AU (1959) Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. *J Comp Physiol Psychol* 52:415–419. [CrossRef Medline](#)
- Martin RF, Bowden DM (2000) Primate brain maps: structure of the macaque brain. Amsterdam: Elsevier Science.
- Maren S, Quirk GJ (2004) Neuronal signalling of fear memory. *Nat Rev Neurosci* 5:844–852. [CrossRef Medline](#)
- Milad MR, Quirk GJ (2012) Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol* 63:129–151. [CrossRef Medline](#)
- Murray EA (2007) The amygdala, reward and emotion. *Trends Cogn Sci* 11:489–497. [CrossRef Medline](#)
- Ochsner KN, Gross JJ (2005) The cognitive control of emotion. *Trends Cogn Sci* 9:242–249. [CrossRef Medline](#)
- Ostroff LE, Cain CK, Bedont J, Monfils MH, Ledoux JE (2010) Fear and safety learning differentially affect synapse size and dendritic translation in the lateral amygdala. *Proc Natl Acad Sci U S A* 107:9418–9423. [CrossRef Medline](#)
- Pape HC, Pare D (2010) Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol Rev* 90:419–463. [CrossRef Medline](#)
- Pare D, Duvarci S (2012) Amygdala microcircuits mediating fear expression and extinction. *Curr Opin Neurobiol* 22:717–723. [CrossRef Medline](#)
- Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439:865–870. [CrossRef Medline](#)
- Paz R, Pare D (2013) Physiological basis for emotional modulation of memory circuits by the amygdala. *Curr Opin Neurobiol* 23:381–386. [CrossRef Medline](#)
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43:897–905. [CrossRef Medline](#)
- Pitman RK, Rasmusson AM, Koenen KC, Shin LM, Orr SP, Gilbertson MW, Milad MR, Liberzon I (2012) Biological studies of post-traumatic stress disorder. *Nat Rev Neurosci* 13:769–787. [CrossRef Medline](#)
- Quirk GJ, Beer JS (2006) Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. *Curr Opin Neurobiol* 16:723–727. [CrossRef Medline](#)
- Rauch SL, Shin LM, Phelps EA (2006) Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research—past, present, and future. *Biol Psychiatry* 60:376–382. [CrossRef Medline](#)
- Rescorla RA (1969) Pavlovian conditioned inhibition. *Psychol Bull* 72:77–94. [CrossRef](#)
- Resnik J, Sobel N, Paz R (2011) Auditory aversive learning increases discrimination thresholds. *Nat Neurosci* 14:791–796. [CrossRef Medline](#)
- Rogan MT, Leon KS, Perez DL, Kandel ER (2005) Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. *Neuron* 46:309–320. [CrossRef Medline](#)
- Saleem K, Logothetis N (2007) A combined MRI and histology atlas of the rhesus monkey brain. London: Elsevier.
- Salzman CD, Fusi S (2010) Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu Rev Neurosci* 33:173–202. [CrossRef Medline](#)
- Salzman CD, Paton JJ, Belova MA, Morrison SE (2007) Flexible neural representations of value in the primate brain. *Ann N Y Acad Sci* 1121:336–354. [CrossRef Medline](#)
- Sangha S, Chadick JZ, Janak PH (2013) Safety encoding in the basal amygdala. *J Neurosci* 33:3744–3751. [CrossRef Medline](#)
- Schechtman E, Laufer O, Paz R (2010) Negative valence widens generalization of learning. *J Neurosci* 30:10460–10464. [CrossRef Medline](#)
- Schiller D, Delgado MR (2010) Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends Cogn Sci* 14:268–276. [CrossRef Medline](#)
- Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008) From fear to safety and back: reversal of fear in the human brain. *J Neurosci* 28:11517–11525. [CrossRef Medline](#)
- Shaban H, Humeau Y, Herry C, Cassasus G, Shigemoto R, Ciochi S, Barbieri S, van der Putten H, Kaupmann K, Bettler B, Lüthi A (2006) Generalization of amygdala LTP and conditioned fear in the absence of presynaptic inhibition. *Nat Neurosci* 9:1028–1035. [CrossRef Medline](#)
- Shabel SJ, Janak PH (2009) Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proc Natl Acad Sci U S A* 106:15031–15036. [CrossRef Medline](#)
- Tye KM, Cone JJ, Schairer WW, Janak PH (2010) Amygdala neural encoding of the absence of reward during extinction. *J Neurosci* 30:116–125. [CrossRef Medline](#)
- van der Kolk BA (1997) The psychobiology of posttraumatic stress disorder. *J Clin Psychiatry* 58:16–24. [CrossRef Medline](#)
- Wiltgen BJ, Silva AJ (2007) Memory for context becomes less specific with time. *Learn Mem* 14:313–317. [CrossRef Medline](#)
- Wimmer GE, Shohamy D (2012) Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science* 338:270–273. [CrossRef Medline](#)
- Xu W, Südhof TC (2013) A neural circuit for memory specificity and generalization. *Science* 339:1290–1295. [CrossRef Medline](#)