

Amygdala-Prefrontal Synchronization Underlies Resistance to Extinction of Aversive Memories

Uri Livneh¹ and Rony Paz^{1,*}

¹Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

*Correspondence: rony.paz@weizmann.ac.il

<http://dx.doi.org/10.1016/j.neuron.2012.05.016>

SUMMARY

Emotional memories can persist for a lifetime but can also undergo extinction. Although we know about the mechanisms involved in expression and extinction, we know very little about the mechanisms that determine whether a specific memory would persist or not. Here, we use partial reinforcement extinction effect (PREE) to explore the neural mechanisms that differentiate persistent from labile memories. We recorded the simultaneous activity of neurons in the amygdala and the dorsal anterior cingulate cortex (dACC) while monkeys engaged in tone-odor aversive conditioning. We report that under continuous reinforcement schedule (ConS), activity in the amygdala precedes behavioral response, whereas under partial schedule (ParS), dACC activity precedes it. Moreover, we find that ParS induced cross-regional pairwise correlations throughout the entire acquisition session, and their magnitude and precision predicted the later resistance to extinction. Our results suggest that memory persistence depends on distributed representations and, specifically, resistance to extinction of aversive memories is maintained by correlated amygdala-dACC activity.

INTRODUCTION

Extinction is a form of inhibitory learning that allows adaptive control of conditioned fear responses (Bouton, 1993; Myers and Davis, 2007). Impaired fear extinction leads to maladaptive and persistent expression of fear in the absence of actual threat and is hypothesized to underlie various mood and anxiety disorders (Delgado et al., 2006; Milad et al., 2006; Myers and Davis, 2007). Physiologically, aberrant activation of plasticity mechanisms at the medial prefrontal cortex (mPFC)-amygdala circuitry (Herry et al., 2010; Herry and Mons, 2004; Muigg et al., 2008; Peters et al., 2010) and sustained activation of neurons that mediate fear expression (Burgos-Robles et al., 2009; Muigg et al., 2008) have been linked to deficits in extinction learning. Yet the contribution of this neural circuitry to the formation of memories that are resistant to extinction remains largely unknown. Specifically, whereas some memories undergo successful extinction, other memories are harder to extinguish and

persist, and the neural mechanisms that differentiate the two are unknown.

To experimentally manipulate resistance to extinction of two otherwise similar aversive memories within the same animal, we took advantage of the behavioral effect of probabilistic reinforcement. Probabilistic schedules can induce slower learning rates, but the effect on the final memory is small (Haselgrove et al., 2004; Leonard, 1975; Rescorla, 1999) and tunable (as shown here). In contrast, the effect on extinction is dramatic and memories that are acquired under probabilistic regime are much harder to extinguish (Haselgrove et al., 2004; Leonard, 1975; Rescorla, 1999). This phenomenon, termed partial reinforcement extinction effect (PREE), provides a unique behavioral tool that can shed light on the neural mechanisms that emerge already during learning and later underlie resistance to extinction. Thus far, although widely used, PREE received little attention as a behavioral tool to explore resistance to extinction of aversive memories.

The amygdala is directly related to enhancement of emotional memories (Hamann et al., 1999; Herry et al., 2008; LeDoux, 2000; Livneh and Paz, 2012; McGaugh, 2004; Pape and Pare, 2010; Paz et al., 2006). The dACC, through its direct connections with the amygdala (Ghashghaei et al., 2007; Pandya et al., 1973; Stefanacci and Amaral, 2002), is thought to regulate expression of learned fear responses (Klavir et al., 2012; Milad et al., 2007), possibly in a similar way to the prelimbic cortex (PL) in rodents (Sierra-Mercado et al., 2011; Vidal-Gonzalez et al., 2006). In addition, the dACC is important for processing of uncertainty (Alexander and Brown, 2011; Rushworth and Behrens, 2008), and human studies suggest differential involvement of dACC during continuous and partial reinforcement schedules (Dunsmoor et al., 2007a; Hartley et al., 2011; Milad et al., 2007). Finally, abnormal functionality of the dACC was observed in anxiety disorders and linked to failure of extinction (Milad et al., 2009; Pannu Hayes et al., 2009; Shin et al., 2011), as in rats with hyperresponsiveness of PL neurons (Burgos-Robles et al., 2009). We therefore hypothesized that differential recruitment of the amygdala-dACC network during acquisition under continuous (ConS) and partial (ParS) schedules could underlie differential resistance to later extinction.

RESULTS

We tested two monkeys on a tone-odor conditioning task (Livneh and Paz, 2010, 2012), with partial reinforcement schedule (ParS) employed in randomly intermingled days with continuous reinforcement schedule (ConS). Each session included a habituation

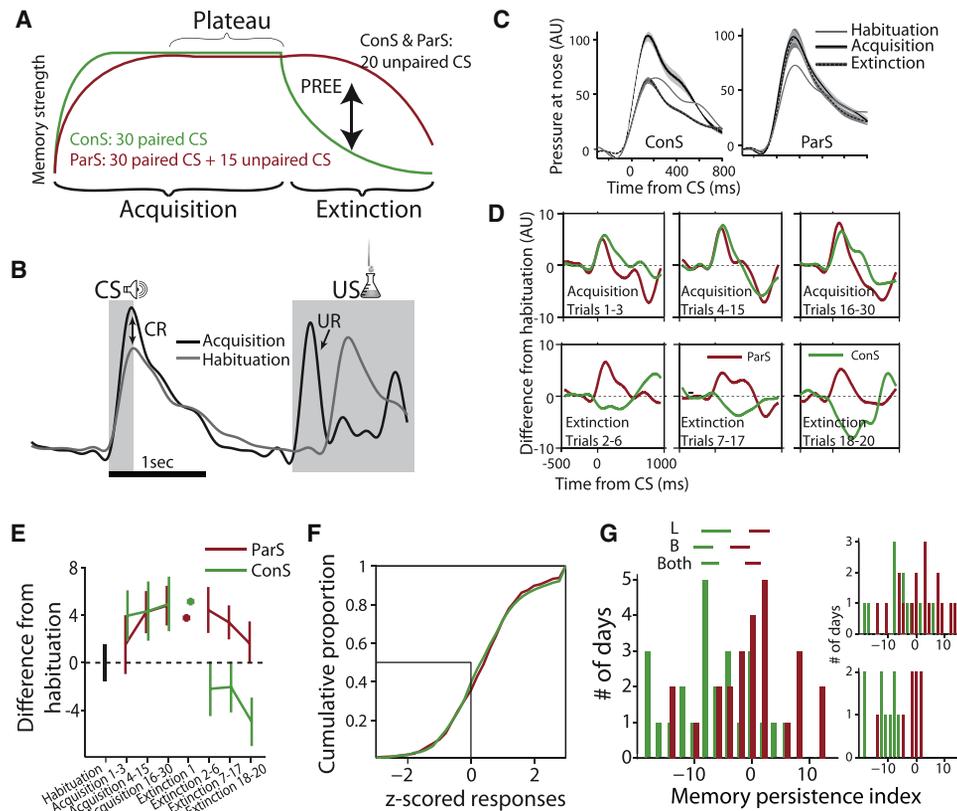


Figure 1. Resistance to Extinction after Partial Reinforcement in a Tone-Odor Conditioning Paradigm

(A) Experimental design: there were two types of sessions (intermingled across the recording period), either continuous reinforcement (ConS) or partial reinforcement (ParS) schedule. Both schedules reach behavioral plateau within a few trials and maintain it for over 25 trials, resulting in a controlled period of memory expression level. However, after acquisition with ConS, extinction training promotes rapid decline of fear, whereas after ParS the decline of fear is delayed.

(B) Trial design: a spontaneous breath onset elicits the conditioned stimulus (CS, a pure tone) and in paired trials, it implies that the onset of the following breath would elicit release of unconditioned stimulus (US, aversive odor). The unconditioned response (UR) was evident by much shorter breaths (inhibition), and the conditioned response (CR) was evident by augmented preparatory breath (real data example is shown).

(C) In sessions with ConS during acquisition (each CS is followed by a US), extinction was fast and complete and CR returned to habituation levels (left). In sessions with ParS (right), extinction was very slow. Mean \pm SEM from two representative sessions are shown.

(D) Each CR was normalized to the response to the tone during the preceding habituation stage, averaged over all available sessions ($n = 25$ for ConS; $n = 25$ for ParS), and presented for different stages during the acquisition and extinction.

(E) CR (quantified as breath volume at 0–350 ms post-CS) was similar during acquisition ($p > 0.1$, interaction, two-way ANOVA) and also in the first trial of extinction ($p > 0.1$, t test) but was significantly different from the second trial of extinction until the end ($p < 0.01$, condition main effect, two-way ANOVA; $p < 0.05$ for all post hoc t tests).

(F) Cumulative distribution of CR during the plateau stage of the acquisition (trials 11–30) indicates that the distribution of CR across trials was similar as well ($p > 0.1$, K-S test).

(G) A memory-persistence index was computed by subtracting CR at the end of extinction (average over trials 11–20) from the CR at the end of acquisition (trials 11–30). The histogram for all sessions reveals a strong effect of condition on memory persistence ($p < 0.001$, t test; $p < 0.05$ for each animal separately, right two histograms). Bars on top of the main histogram indicate mean \pm SEM of the main distribution and for each animal separately.

stage (unpaired presentations of a tone, the CS), an acquisition stage (30 paired presentations of the CS followed by an aversive odor, the US, in a trace-conditioning paradigm), and an extinction stage (unpaired CS). In ParS sessions, trials were reinforced in a 2:1 ratio, with keeping the overall number of reinforced trials equivalent for ParS and ConS. In preliminary sessions, we tuned the reinforcement ratio to yield minimal difference between memory expressions at the end of learning. The aim was to obtain a plateau period at the end of learning, in which memory expression levels are similar in ParS and ConS sessions, but when similar extinction training later on would yield differential

results. Hence, in this controlled plateau period, although behavior appears similar, the underlying mechanisms of persistent and more labile memories should be different (Figure 1A).

Partial Reinforcement Impairs Extinction of Aversive Tone-Odor Associations

Memory expression level was measured by the modulation in the volume of the breath that follows the tone but precedes odor release (Figures 1B and 1C). Such preparatory breath modulation was apparent already after three acquisition trials (Figures 1D and 1E, $p < 0.01$, ANOVA). In line with our design, although

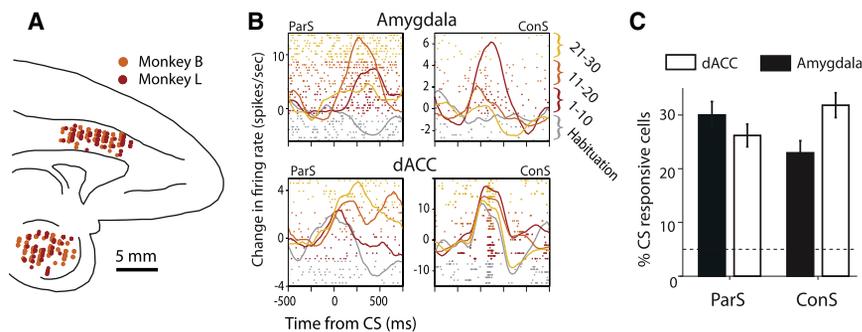


Figure 2. Recording Sites and Single Neuron Responses to the CS

(A) Single neurons were recorded simultaneously from the right amygdala ($n = 131$) and dACC ($n = 172$). Recording sites were reconstructed based on MRI with calibrating electrodes (Figure S2) and shown here with lateral projection on a midbrain scheme of the macaque brain. (B) Single cell examples of different types of responses to the CS during habituation (gray) and the three stages of the acquisition (red-yellow shades). Mean change in firing rate (PSTH) overlaid on raster plots. First row, two amygdala neurons; second row, two dACC neurons.

(C) Cells were classified responsive if their tone-evoked activity during acquisition was significantly different than their tone-evoked activity during habituation. The number of overall responsive cells was not different across conditions and across regions ($p > 0.1$, interactions and main effects, two-way ANOVA).

ParS had a slightly slower learning rate (Figure 1E, but not significantly, $p > 0.3$, ANOVA), we observed similar expression levels in ParS and ConS during the plateau phase (Figure 1C, trials 4–30, $p > 0.5$, condition main effect, two-way ANOVA), and these were also similarly distributed across trials (Figure 1F, $p > 0.1$, K-S test). We also verified that the magnitude of the UR was comparable, suggesting that the monkeys perceived the odor to be similarly aversive in both conditions ($p > 0.5$, see Figures S1A and S1B available online).

In contrast and as expected from previous work on PREE, extinction trials that followed ConS dropped already at the second trial but remained high throughout the extinction training that followed ParS acquisition ($p < 0.001$, two-way ANOVA interaction effect; Figures 1C–1E). We computed a memory-persistence index by subtracting the memory expression level at the end of extinction (trials 11–20) from that at the end of acquisition when behavior is at plateau (trials 11–30). Memory persistence differed significantly across conditions and also when tested separately for each animal ($p < 0.05$ for all comparisons, Figure 1G). To further validate that this is due to the acquisition regime, we compared each CR to baseline inhales from the intertrial intervals. There was no difference between the session types during habituation ($p > 0.1$, t test, Figure S1C), but CRs returned to baseline during ConS extinction and remained elevated during ParS extinction ($p < 0.01$, two-way ANOVA).

To conclude, although acquisition reached a similar plateau of expression level, extinction was fast under ConS and very slow under ParS (ranging from slow to none, as seen in the distribution in Figure 1G).

Dissociation of Amygdala and dACC Roles in Mediating Learned Fear Responses

To evaluate interactions in the amygdala-dACC pathway under both conditions, we simultaneously recorded single-unit activity from both regions (Figure 2A; amygdala: $n = 131$; dACC: $n = 172$; Figures S2A and S2B, Table S1). Neural responses to the CS were normalized and compared against tone responses at habituation (Figure 2B), revealing that 26% of amygdala neurons and 29% of dACC neurons had significant acquired responses (both higher than chance, $p < 0.001$, χ^2), and there was no interaction or main effect of schedule or region (Figure 2C, $p > 0.1$ for all,

two-way ANOVA). In both the amygdala and the dACC, responsive cells were homogeneously distributed within our recording borders (Figure S2C, $p > 0.2$ for all, bootstrap analysis), suggesting that they represent an activity pattern common in wide parts of these two structures. In addition, there was no effect of reinforcement schedule on neural responses to the US (Figures S2D and S2E, $p > 0.1$, two-way ANOVA).

We then inspected the temporal relationship between neuronal and behavioral responses. To do so, we computed trial-by-trial cross-correlations between the firing rate (FR) and the breathing response at all delays from the CS (Figure S3A). Significant bins above the diagonal in such a correlation matrix indicate that changes in FR precede the behavioral response and significant bins below the diagonal indicate that changes in FR follow behavior. Although the overall number of significant bins was not different between regions and schedules ($p > 0.1$, two-way ANOVA, Figure S3B), inspection of individual matrices revealed that amygdala neurons were more likely to fire before behavior under ConS, whereas dACC neurons were more likely to fire before behavior under ParS (Figure 3A), as also shown by the proportions of significant bins above (Figure 3B, top) and below (bottom) the diagonal (Figure 3B, $p < 0.001$, three-way ANOVA, Figure S3C). To further validate this, we computed the center of mass of the correlation for all neurons and its distance from the diagonal, which provides a better estimation for directionality because it takes into account the strength of the correlations as well. This analysis indicated that amygdala activity indeed precedes behavior under ConS, and dACC activity precedes behavior under ParS (Figure 3C, $p < 0.01$, two-way ANOVA, interaction of distance from diagonal with brain region and schedule as factors, confirmed by post hoc comparisons). Finally, we computed a directionality index (sum of correlations above the diagonal minus sum of correlations below the diagonal divided by total sum of all correlations), revealing again the same pattern of temporal relationship between neural activity and behavior (Figure 3D; $p < 0.01$, two-way ANOVA). For robustness, we also repeated all analyses using bootstrap methods for bin-wise significance (Figures S3D and S3E) (Paz et al., 2006, 2009).

We conclude that there is a double dissociation in the roles of the amygdala and the dACC in mediating acquired aversive responses for partial and continuous reinforcement schedules.

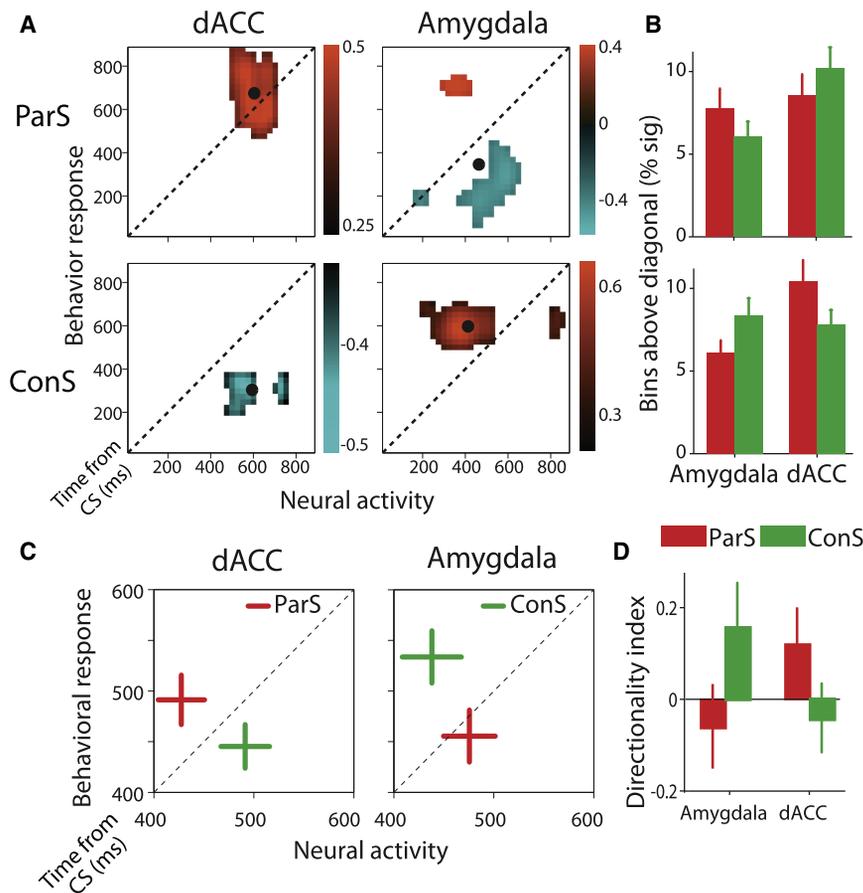


Figure 3. Double Dissociation between Behavior and Firing Rates for the Amygdala and dACC during ConS and ParS

(A) Correlation matrices between the neural and behavioral responses at different time lags from the CS. Correlations show that neural activity could either precede the behavioral response, if most correlations are above the main diagonal (top left and bottom right matrices), or follow behavioral response, if most correlations are below the main diagonal (top right and bottom left). Typical examples from dACC (left column) and amygdala (right column) during ParS (top row) or ConS (bottom row) are shown, with positive significant correlations in orange shades and negative correlations in light blue shades. The center of mass is indicated by a thick black dot. See Figures S3A–S3C for construction of such correlations, full matrices, and population summary.

(B) Percentage of significant bins above the main diagonal (top) and below the main diagonal (bottom) ($p < 0.001$, three-way ANOVA).

(C) The location of center of mass reveals the double dissociation: dACC responses precede behavior during ParS, and amygdala responses precede behavior during ConS. Crosses are mean \pm SEM of center of mass from all sessions.

(D) Directionality index (“sum of correlations above the diagonal” minus “sum of correlations below the diagonal” divided by “total sum of all correlations”) further supports the double dissociation, indicating positive direction of dACC and amygdala responses for ParS and ConS, respectively ($p < 0.01$, two-way ANOVA and post hoc comparisons).

Synchronized Amygdala-dACC Activity Maintains Memory during ParS

We next turned to examine the correlations between simultaneously recorded pairs of amygdala and dACC neurons ($n = 483$ pairs) and tested the hypothesis that altered correlation patterns at acquisition underlie the delayed extinction. Correlation matrices at all delays (in 25 ms bins) were computed for each of ten trials from the habituation and the acquisition phases and significance was assessed per bin (Figures 4 and S4) (Paz et al., 2006). The mean number of significant bins was then normalized by the distribution of significant bins in the shuffled data of the corresponding pair, producing a correlation-factor index. This did not differ during habituation between ConS and ParS ($p > 0.3$, t test). During acquisition, we found that ConS induced an early sharp increase in correlations that diminished to habituation levels in parallel to reaching plateau level of behavior. In contrast, ParS induced an increase that remained high throughout the whole acquisition stage, even much after behavioral plateau was achieved (Figures 5A and 5B), and this dynamic was evident in the correlation-factor index (Figure 5C, $p < 0.001$; interaction in two-way ANOVA). When the correlation factor was inspected separately above and below the diagonal, the same effects were observed (Figure 5C, insets, $p < 0.05$, interaction in two-way ANOVA), suggesting that it is an overall increase in reciprocal interactions between the two regions. For robustness, we repeated the analyses using more conven-

tional statistical tests to assess bin significance (Pearson statistics) and revealed the same dynamics (Figure S5A, $p < 0.001$ in two-way ANOVA). Notice that analyses are performed only on reinforced trials; hence, the number of preceding reinforced trials is identical for each data point in the comparison between ParS and ConS. Finally, in accordance with the distribution of projections from dACC (area 24) to the amygdala (Stefanacci and Amaral, 2002), we found more significant correlations at medial penetrations to the amygdala than at lateral ones (Figure S5B; $p < 0.01$, main effect in two-way ANOVA).

Amygdala-dACC Synchrony Predicts Later Resistance to Extinction

If the sustained correlations at the end of the acquisition indeed support a more resistant memory, then their magnitude should somehow predict the time it takes to extinguish. To test this, we compared the mean correlation-factor index with the memory-persistence index (as in Figure 1G) on an individual session basis. We found a significant positive relationship indicating that increased amygdala-dACC correlations at the end of acquisition predict memory persistence (Figure 6A, $r = 0.61$, $p < 0.0002$). This correlation was mostly due to ParS days (Figure 6A, top inset, $r = 0.72$; $p < 0.002$), in which memory persistence was observed, and with marginal contribution from ConS days (bottom inset, $r = 0.46$; $p = 0.05$) and was replicated when the correlation-factor index was obtained using standard

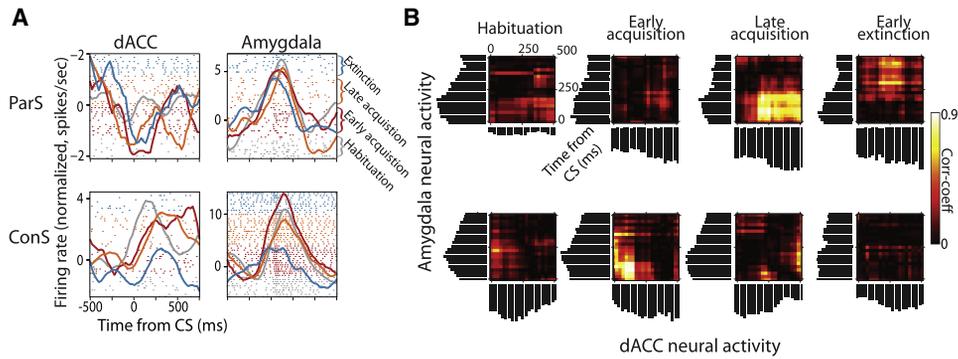


Figure 4. Amygdala-dACC Correlations during Acquisition: Pair Examples

(A) Pairs of amygdala-dACC neurons were recorded simultaneously at either ParS (top pair) or ConS (bottom pair) sessions during habituation (gray), acquisition (red and orange), and early extinction (blue).

(B) Correlation matrices between the cells shown in (A) revealed enhanced correlation during late acquisition at ParS sessions, whereas in ConS sessions, correlation was enhanced during early acquisition. Much smaller correlations were observed at either habituation or extinction. Corresponding PSTHs of the dACC neuron (x axis) and the amygdala neuron (y axis) are shown.

Pearson statistics (Figure S6A) and for each animal separately (Figures S6C and S6D). This result cannot be attributed to a higher number of co-occurrence of responsive cells in the two regions because of the following: (1) we used shuffling techniques to correct for this possibility; (2) it is not predicted by the pattern of individual responsive cells (Figure S6B); and (3) moreover, local intraregional pairwise correlations of either amygdala-amygdala pairs (Figure 6B, top row) or dACC-dACC pairs (Figure 6B, bottom row) failed to differentiate the two conditions and failed to explain the behavioral resistance to extinction.

Finally, we found that synchronized activity became more adjacent (locked) to the CS during ParS (Figure 6C, $p < 0.001$,

interaction in two-way ANOVA confirmed by post hoc, $p < 0.05$), and the distance of the center of mass was negatively correlated with memory persistence on ParS days (Figure 6D, $r = -0.62$; $p < 0.01$), suggesting that the timing of the correlations also contributes to memory persistence.

DISCUSSION

In this study, we demonstrate that resistance to extinction of aversive memories can depend on the neural mechanisms that are activated already when the memory is formed. Using partial (ParS) and continuous (ConS) reinforcement, we were able to

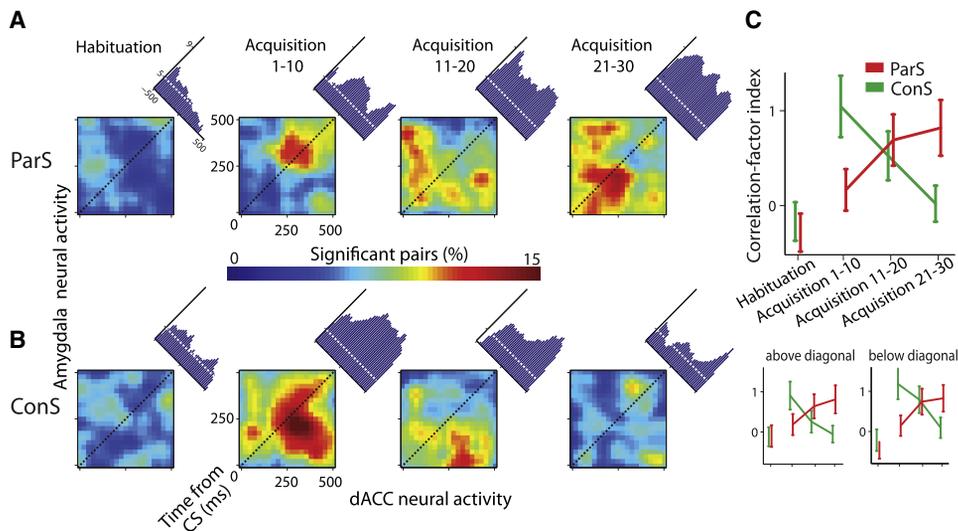


Figure 5. Amygdala-dACC Correlations during Acquisition

Correlation matrices were computed for all available pairs of simultaneously recorded dACC and amygdala neurons ($n = 483$) in different time lags from the CS. (A and B) Color-coded maps show the proportion of pairs with significant correlation at each bin in the different stages of the learning paradigm for ParS (A) and ConS (B). The corresponding cross-correlation, with the chance-level proportion marked by a dashed line, is shown at the top right of each panel.

(C) During ConS, correlations sharply increased and then dropped back to baseline. In contrast, during ParS, correlations increased gradually and remained high throughout the acquisition, even much after behavior reached plateau. These dynamics are the same when considering separately the correlations above the diagonal (bottom left inset) and below it (bottom right inset). See text for statistics.

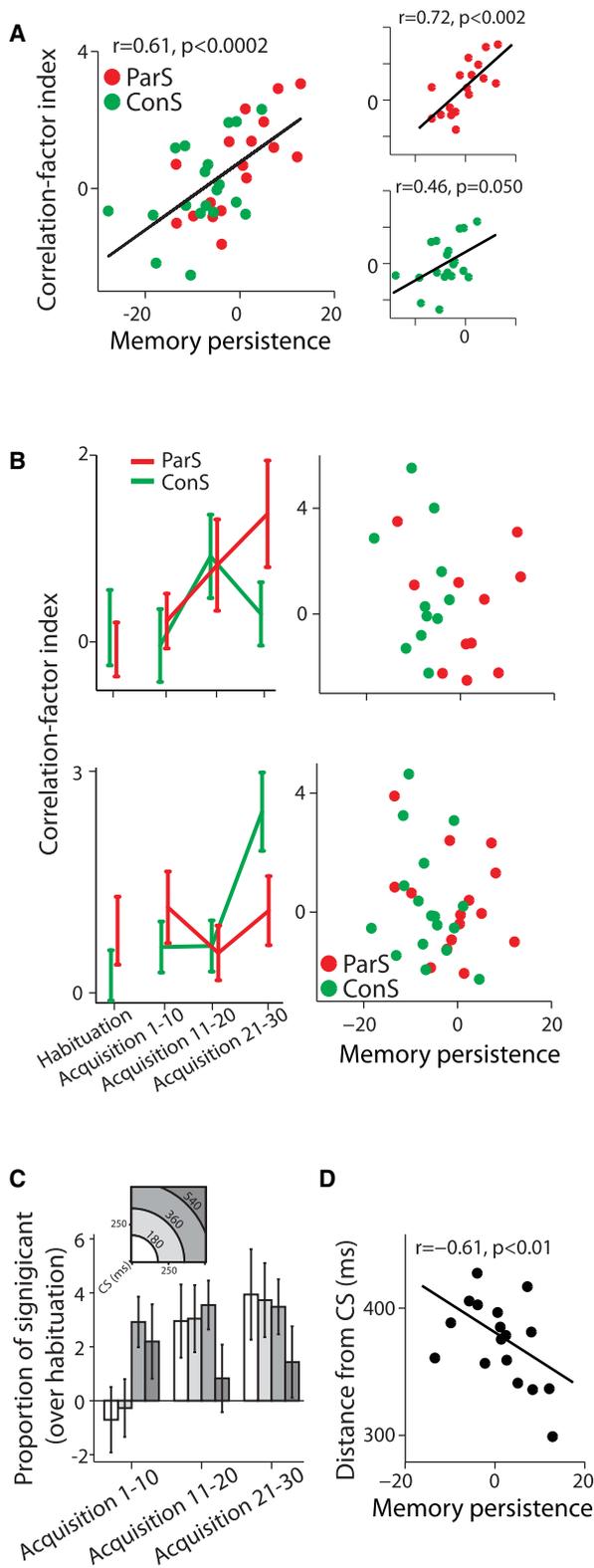


Figure 6. Amygdala-dACC Synchronization during Acquisition Predicts Later Resistance to Extinction

(A) Amygdala-dACC mean correlation-factor index at the end of each acquisition session positively correlated with resistance to extinction in that session,

repeatedly create two types of memory in the same animal (but in different randomly alternating days)—one that undergoes extinction within a few trials and one that undergoes extinction much more slowly. This persistent expression of fear (resistance to extinction) after ParS training implies that the tone-odor associations were acquired differently under the different schedules. We verified that memory expression levels were similar and evenly distributed across trials before extinction started, giving us a controlled time period in which we could observe these differential mechanisms. We find that correlated and synchronized amygdala-prefrontal activity supports and maintains the memory under ParS condition, and the magnitude and precision of these correlations reliably predict the later resistance to extinction.

The amygdala is thought to be sufficient for acquisition of simple fear associations, but it participates in mediating more complex emotional memories as well (LeDoux, 2000; McGaugh, 2004). Direct sensory inputs that converge on single cells in the amygdala and induce synaptic plasticity have been directly related to acquisition of fear (Herry et al., 2008; Pape and Pare, 2010; Paré et al., 2004), and the rich repertoire of inputs to the primate amygdala suggests that it is highly adapted for rapid detection of sensory associations (Baxter and Murray, 2002; Ghashghaee et al., 2007; McDonald, 1998; Stefanacci and Amaral, 2002). However, it is less clear to what extent the amygdala alone can support complex forms of learning (Bryden et al., 2011; Holland and Gallagher, 2004; Li et al., 2011; Roesch et al., 2010; Vazdarjanova and McGaugh, 1998) and specifically probabilistic relationships as in partial reinforcement. The dACC has been implicated with monitoring of behavior, attention, signaling of error likelihood, and reinforcement volatility (Carter et al., 1998; Rushworth and Behrens, 2008; Wallis and Kennerley, 2010) and can therefore be more adept for learning complex relationships and contingencies. This is in line with our finding that

as measured by the persistence index (as in Figure 1G). This was so when pooling ParS and ConS ($p < 0.0002, r = 0.61$) and for ParS alone ($p < 0.002, r = 0.72$, top right inset) and less so for ConS ($p = 0.05, r = 0.46$, bottom right inset). Notice this is a real prediction, because correlations were taken from the end of the acquisition, before extinction begins.

(B) Intraregional correlations were computed in the same way as for interregional correlations for all available pairs of amygdala-amygdala neurons ($n = 159$, top row) and dACC-dACC neurons ($n = 252$, bottom row) that were recorded simultaneously on separate electrodes. Although interesting dynamics are observed, it did not differ across the reinforcement schedules (left column; $p > 0.2$ for main effects and interaction, two-way ANOVA), and there were no correlations between the proportion of significant bins and memory persistence at either the amygdala-amygdala pairs or dACC-dACC pairs (right columns, $p > 0.1$ for all Pearson coefficients, for each schedule and region separately and combined).

(C) Interregional correlation matrices were divided into arched subdivisions with equal distances from the CS (origin of axes), as shown in the inset. The proportion of significant bins was calculated for each subdivision and each phase of the acquisition. In ParS days, this proportion increases with acquisition (three-way ANOVA interaction, $p < 0.0001$, confirmed by post hoc comparisons), whereas in ConS days, proportion in all arches diminishes gradually (data not shown).

(D) Distance of the center of mass of the correlation matrices from the CS presentation negatively correlates and predicts the persistence index for ParS sessions ($p < 0.01, r = -0.62$).

activity in the dACC precedes the behavioral response at ParS, whereas neural activity in the amygdala precedes behavior during ConS, although the behavioral response itself was indistinguishable in both conditions.

The synchronized discharge of both regions spiked at the beginning of learning but dropped back to baseline within a few trials of ConS. One option is that amygdala-dACC interactions are required for the initial learning phase, but not for the maintenance of the memory once it is formed and synaptic changes are made downstream. Another option is that the dACC takes an active part by default but then lowers its communication with the amygdala when it realizes that it is not required for the simple associations. This can be achieved by feedback reports about correct behavior. In sharp contrast to ConS, the amygdala-dACC synchronized activity maintained during ParS, even much after behavioral plateau was obtained and was similar for ConS and ParS (trials 4–30). This finding suggests that these correlations are required for active maintenance of the memory under ParS. This is further supported by the fact that the magnitude of these correlations at the end of learning, and their locking to CS, were a reliable predictor for the difficulty (length) of the following extinction training.

Why should amygdala-dACC correlations make the memory harder to extinguish? Extinction is a new learning that was shown to be mediated by subregions of the medial prefrontal cortex (mPFC). This includes the rodent infralimbic cortex (IL) (Milad and Quirk, 2002; Sierra-Mercado et al., 2011) and the primate vmPFC (Phelps et al., 2004). These regions exhibit opposite activation patterns to that of the amygdala and are activated during extinction recall, whereas the amygdala is inhibited. The primate dACC was shown to have the opposite effect on fear expression and extinction (Dunsmoor et al., 2007b; Milad et al., 2007), similar to the rodent prelimbic cortex (PL) (Sierra-Mercado et al., 2011; Vidal-Gonzalez et al., 2006), and promotes fear in general (Burgos-Robles et al., 2009). Hence, these are probably two competing pathways with opposite effects. If the memory is maintained primarily in synaptic changes in the amygdala and further downstream (Ciocchi et al., 2010; Duvarci et al., 2011), then the vmPFC pathway would have an easier job of inhibiting it. However, if the memory is actively maintained by the amygdala-dACC pathway, then the vmPFC pathway would have a much harder job and it would take longer to “undo.” In addition, prolonged and enhanced interregional correlations could strengthen synaptic mechanisms and plasticity and induce cellular and molecular changes that were described in this timeframe of dozens of minutes (our acquisition stage lasts for about 30 min). Complementing this, increased coupling between amygdala and/or hippocampal prefrontal circuits has been shown to parallel differences in extinction and consolidation of emotional memories (Adhikari et al., 2010; Lesting et al., 2011; Narayanan et al., 2011; Paz et al., 2007; Popa et al., 2010; Sangha et al., 2009).

It was recently shown that there is a shift of balance between the amygdala and the mPFC for learning of extinction versus its relearning. Specifically, learning to inhibit fear for the first time requires NMDA receptors in the amygdala (Laurent et al., 2008), whereas relearning extinction involves NMDA receptors in the mPFC (Laurent and Westbrook, 2008). Our paradigm

involves daily acquisition and extinction of aversive memories, and hence all of our experiment was conducted in a relearning scenario. Nevertheless, we were able to continuously obtain a difference between ConS and ParS sessions along the whole recording period, and we verified that our main results (resistance to extinction in ParS and fast extinction in ConS, as in Figure S1C; the dissociation between early and late acquisition in amygdala-dACC neural correlations, as in Figure 5C; and the prediction of resistance to extinction by cross-regional correlations, as in Figure 6A) were significant when tested separately for early recording days (the first half) and for late recording days (the second half) and were not significantly different between early and late sessions. An interesting possibility therefore is that the distinction between first-time learning and relearning applies to the difference between ConS and ParS. Due to the uncertainty of the CS-US contingency, the association might need to be relearned within a session, and hence the mPFC might be more involved during ParS, as was indeed observed here. Humans are usually well experienced with anxiety-evoking stimuli and with emotional regulation of it. From this perspective, relearning of fear and its extinction might be an adequate model for anxiety-related disorders.

Indeed, unlike naive rats used in many studies, human patients are almost always exposed to the stimulus before it becomes associated with fear (e.g., the twin towers as a workplace before 9/11, the personal car before the crash, etc.). These exposures can be thought of as unreinforced trials. When such a context or stimulus suddenly and surprisingly becomes unsafe, much uncertainty is involved and the dACC plays a more active role. According to this model, early extinction might be similar to late acquisition under ParS, because they cannot be distinguished statistically. This might suggest that reducing the strength of dACC outputs to the amygdala or related structures during early extinction may improve efficacy of the extinction process (or reduce efficacy of the original fear memory) and perhaps prevent later spontaneous recovery. This is in line with recent results from our laboratory, showing that low-frequency stimulation in the dACC of monkeys during extinction learning can depress the region and diminish spontaneous recovery of aversive associations when measured 24 hr later (Klavriv et al., 2012).

Although several studies have explored variability across animals in the extinction process itself, little is known about the neural changes that occur already during acquisition and that could make a specific memory more resistant to extinction. Here we describe one such mechanism. Full (100%) contingencies are rare in real life and partial reinforcement could therefore serve as a realistic model for anxiety disorders and PTSD and improve translatability (Milad and Quirk, 2012). Surprise and attention signals were identified in single neurons of the amygdala (Belova et al., 2007; Li et al., 2011; Roesch et al., 2010) and the dACC (Bryden et al., 2011; Hayden et al., 2011) (albeit with different characteristics). Such signals occur during partial reinforcement and can initiate and maintain the sustained synchronized activity across the two regions as we describe here. This underlying mechanism could in turn make the aversive memory more resistant to extinction, as observed in clinical cases.

EXPERIMENTAL PROCEDURES

Animals

Two male macaca fascicularis (4–7 kg) were implanted with a recording chamber (27 × 27 mm) above the right amygdala and dACC under full 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.

MRI-Based Electrode Positioning

Anatomical 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 three-dimensional (3D) gradient-echo (MPRAGE) pulse sequence was acquired with repetition time (TR) of 2,500 ms, echo time (TE) of 3.36 ms, 8° flip angle, and two averages. Images were acquired in the sagittal plane, 192 × 192 matrix and 0.8³ mm or 0.6³ mm resolution. A scan was performed before surgery and used to align and guide the positioning of the chamber on the skull for each individual animal (by relative location of the amygdala and anatomical markers of the interaural line and the anterior commissure). After surgery, we performed another scan with two electrodes directed toward the amygdala and the dACC, and two to three observers separately inspected the images and calculated the anterior-posterior and lateral-medial borders of the amygdala and dACC relative to each of the electrode penetrations. The depth of the regions was calculated from the dura surface.

Recordings

Each day, three to six 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 at 2–5 mm in the cortex. Electrodes were then moved independently into the amygdala and the dACC (we performed four to seven 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 band-pass filtered, and sampled at 25 kHz, and online 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, offline spike sorting was performed for all sessions to improve unit isolation (offline sorter, Plexon).

Behavior

Monkeys were seated in a chair with a custom-made nasal mask attached to their nose (Livneh and Paz, 2010). The mask was attached to two pressure sensors with different sensitivity range (1/4" and 1" H₂O pressure range, AllSensors) that enable real-time detection of breath onset. Experimental sessions initiated by a habituation session of ten presentations of the CS (a pure tone chosen randomly from 1,000–2,400 Hz, delivered through an Adam5 speaker, ADAM Audio GmbH). The acquisition session that followed included 30 trials of CS paired with an aversive odor (3 s; 1:20 solution of propionic acid distilled in mineral oil; Sigma-Aldrich). Propionic acid stimulates olfactory and trigeminal receptors at the nose and is highly aversive to humans and monkeys. CS was triggered by breath onsets, and odor (US) was released at the following breath onset (but not before 1 s elapsed). On ParS days, an additional 15 presentations of unpaired CS were intermingled with the paired CSs; hence, the overall number of reinforced trials was equal in ParS and ConS days. In ConS days, sham trials (neither CS nor US) were implanted into the paradigm to maintain equal total length of the acquisition stage. Twenty unpaired CSs were presented to the monkey in order to extinguish the acquired association between the CS and the US. We used immediate extinction because spontaneous recovery is evident after immediate extinction, indicating that the memory is inhibited rather than erased. Breath volumes were

computed as the area under the curve 0–350 ms from breath onset and normalized by the habituation breath volume. US, i.e., responses to the odor (Figure S1), were computed 100–900 ms postodor release.

Data Analysis

Firing rates (FR) were computed with a rectangular 300 ms window that was advanced at 25 ms steps. We computed the neuron-evoked responses for the CS and the US at nonoverlapping bins during 0–500 ms post-CS. We computed the spontaneous breath-evoked response of the neurons at habituation and acquisition and stimuli-evoked responses were normalized by subtracting the mean spontaneous breath-evoked responses and dividing by the standard deviation. Stimuli-evoked responses were identified as significant if they significantly change and increase the magnitude of their response to the CS from the habituation response ($p < 0.05$, two-tail Wilcoxon test). Only neurons that fired in at least five out of the 40 trials (30 acquisition, 10 habituation) were included for further analyses (Table S1).

To evaluate correlation between single neuron activity and the behavioral response, we correlated the FR matrix of the neuron response at 0–900 ms with the matrix of momentary pressure at the nose as 0–900 ms post-CS presentations using the following equation:

$$\text{Corr}(FR_i, BRT_j) = \frac{\text{Cov}(FR_i, BRT_j)}{\sigma_i \sigma_j}$$

where FR_i is the firing rate of the neuron at time i post-CS, and BRT_j is pressure at the nose at time j post-CS. Bins of the correlation matrix were identified as significant at $p < 0.05$ and only if significant bins formed at least a 2 × 2 cluster (actual p value is therefore $0.05^4 = 0.0000625$). We further characterized the center of mass of significant bins in each matrix and its distance from the diagonal. In addition, directionality index, di , was computed as $di = (a-b)/(a+b)$, where a is the squared summed correlation above the diagonal and b is the squared summed correlation below the diagonal. di range was between -1 and 1 , where 1 indicates dominance of correlations that precede the behavioral response and -1 indicates dominance of correlations that followed the behavioral response.

To evaluate neuronal interaction between dACC and amygdala neurons, we computed correlation matrices between all available pairs of dACC and amygdala neurons. Correlations were computed for each of ten trials at habituation and acquisition. We used shuffling techniques to assess statistical significance (Aertsen et al., 1989; Paz et al., 2006, 2009). We shuffled the order of the trials 100 times and computed for each bin the distribution of the squared correlation (i.e., the percentage of explained variance) at the shuffled condition. Bins with squared correlation that exceeded 95% of the shuffled distribution were identified as significant, but only if they were part of a cluster of at least 2 × 2 significant bins. Similar results were obtained when conventional significance tests of correlation were employed. Lastly, to assess the overall strength of each correlation matrix, the overall number of significant bins was normalized by the mean and the standard deviation of the significant bins in the shuffled matrices, creating unbiased correlation-factor index that can be averaged across multiple matrices (i.e., pairs of amygdala-dACC neurons).

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.neuron.2012.05.016>.

ACKNOWLEDGMENTS

We thank Yossi Shohat for invaluable contribution to the work and welfare of the animals; Dr. Gil Hecht, Dr. Eilat Kahana, and Dr. Gal Marjan for help with medical and surgical procedures; and Dr. Edna Furman-Haran and Nachum Stern for MRI procedures. This work was supported by NIP1 2010-11-b5, ISF 430/08, and ERC-FP7-StG 281171 grants to R.P.

Accepted: May 8, 2012
Published: July 11, 2012

REFERENCES

- Adhikari, A., Topiwala, M.A., and Gordon, J.A. (2010). Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron* 65, 257–269.
- Aertsen, A.M., Gerstein, G.L., Habib, M.K., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *J. Neurophysiol.* 61, 900–917.
- Alexander, W.H., and Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344.
- Baxter, M.G., and Murray, E.A. (2002). The amygdala and reward. *Nat. Rev. Neurosci.* 3, 563–573.
- Belova, M.A., Paton, J.J., Morrison, S.E., and Salzman, C.D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55, 970–984.
- Bouton, M.E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol. Bull.* 114, 80–99.
- Bryden, D.W., Johnson, E.E., Tobia, S.C., Kashtelyan, V., and Roesch, M.R. (2011). Attention for learning signals in anterior cingulate cortex. *J. Neurosci.* 31, 18266–18274.
- Burgos-Robles, A., Vidal-Gonzalez, I., and Quirk, G.J. (2009). Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *J. Neurosci.* 29, 8474–8482.
- Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747–749.
- Ciocchi, S., Herry, C., Grenier, F., Wolff, S.B., Letzkus, J.J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M.B., et al. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* 468, 277–282.
- Delgado, M.R., Olsson, A., and Phelps, E.A. (2006). Extending animal models of fear conditioning to humans. *Biol. Psychol.* 73, 39–48.
- Dunsmoor, J.E., Bandettini, P.A., and Knight, D.C. (2007a). Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. *Behav. Neurosci.* 121, 635–642.
- Dunsmoor, J.E., Bandettini, P.A., and Knight, D.C. (2007b). Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. *Behav. Neurosci.* 121, 635–642.
- Duvarci, S., Popa, D., and Paré, D. (2011). Central amygdala activity during fear conditioning. *J. Neurosci.* 31, 289–294.
- Ghashghaei, H.T., Hilgetag, C.C., and Barbas, H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage* 34, 905–923.
- Hamann, S.B., Ely, T.D., Grafton, S.T., and Kilts, C.D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nat. Neurosci.* 2, 289–293.
- Hartley, C.A., Fischl, B., and Phelps, E.A. (2011). Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. *Cereb. Cortex* 21, 1954–1962.
- Haselgrove, M., Aydin, A., and Pearce, J.M. (2004). A partial reinforcement extinction effect despite equal rates of reinforcement during Pavlovian conditioning. *J. Exp. Psychol. Anim. Behav. Process.* 30, 240–250.
- Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., and Platt, M.L. (2011). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J. Neurosci.* 31, 4178–4187.
- Herry, C., and Mons, N. (2004). Resistance to extinction is associated with impaired immediate early gene induction in medial prefrontal cortex and amygdala. *Eur. J. Neurosci.* 20, 781–790.
- Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., and Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature* 454, 600–606.
- Herry, C., Ferraguti, F., Singewald, N., Letzkus, J.J., Ehrlich, I., and Lüthi, A. (2010). Neuronal circuits of fear extinction. *Eur. J. Neurosci.* 31, 599–612.
- Holland, P.C., and Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Curr. Opin. Neurobiol.* 14, 148–155.
- Klavriv, O., Genuit-Gabai, R., and Paz, R. (2012). Low-frequency stimulation depresses the primate anterior-cingulate-cortex and prevents spontaneous recovery of aversive memories. *J. Neurosci.* Published online June 20, 2012. <http://dx.doi.org/10.1523/JNEUROSCI.6481-11.2012>.
- Laurent, V., and Westbrook, R.F. (2008). Distinct contributions of the basolateral amygdala and the medial prefrontal cortex to learning and relearning extinction of context conditioned fear. *Learn. Mem.* 15, 657–666.
- Laurent, V., Marchand, A.R., and Westbrook, R.F. (2008). The basolateral amygdala is necessary for learning but not relearning extinction of context conditioned fear. *Learn. Mem.* 15, 304–314.
- LeDoux, J.E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184.
- Leonard, D.W. (1975). Partial reinforcement effects in classical aversive conditioning in rabbits and human beings. *J. Comp. Physiol. Psychol.* 88, 596–608.
- Lesting, J., Narayanan, R.T., Kluge, C., Sangha, S., Seidenbecher, T., and Pape, H.C. (2011). Patterns of coupled theta activity in amygdala-hippocampal-prefrontal cortical circuits during fear extinction. *PLoS ONE* 6, e21714.
- Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A., and Daw, N.D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nat. Neurosci.* 14, 1250–1252.
- Livneh, U., and Paz, R. (2010). An implicit measure of olfactory performance for non-human primates reveals aversive and pleasant odor conditioning. *J. Neurosci. Methods* 192, 90–95.
- Livneh, U., and Paz, R. (2012). Aversive-bias and stage-selectivity in neurons of the primate amygdala during acquisition, extinction and overnight retention. *J. Neurosci.* Published online June 20, 2012. <http://dx.doi.org/10.1523/JNEUROSCI.0323-12.2012>.
- McDonald, A.J. (1998). Cortical pathways to the mammalian amygdala. *Prog. Neurobiol.* 55, 257–332.
- McGaugh, J.L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.* 27, 1–28.
- Milad, M.R., and Quirk, G.J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* 420, 70–74.
- Milad, M.R., and Quirk, G.J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. *Annu. Rev. Psychol.* 63, 129–151.
- Milad, M.R., Rauch, S.L., Pitman, R.K., and Quirk, G.J. (2006). Fear extinction in rats: implications for human brain imaging and anxiety disorders. *Biol. Psychol.* 73, 61–71.
- Milad, M.R., Quirk, G.J., Pitman, R.K., Orr, S.P., Fischl, B., and Rauch, S.L. (2007). A role for the human dorsal anterior cingulate cortex in fear expression. *Biol. Psychiatry* 62, 1191–1194.
- Milad, M.R., Pitman, R.K., Ellis, C.B., Gold, A.L., Shin, L.M., Lasko, N.B., Zeidan, M.A., Handwerker, K., Orr, S.P., and Rauch, S.L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol. Psychiatry* 66, 1075–1082.
- Muigg, P., Hetzenauer, A., Hauer, G., Hauschild, M., Gaburro, S., Frank, E., Landgraf, R., and Singewald, N. (2008). Impaired extinction of learned fear in rats selectively bred for high anxiety—evidence of altered neuronal processing in prefrontal-amygdala pathways. *Eur. J. Neurosci.* 28, 2299–2309.
- Myers, K.M., and Davis, M. (2007). Mechanisms of fear extinction. *Mol. Psychiatry* 12, 120–150.
- Narayanan, V., Heiming, R.S., Jansen, F., Lesting, J., Sachser, N., Pape, H.C., and Seidenbecher, T. (2011). Social defeat: impact on fear extinction and amygdala-prefrontal cortical theta synchrony in 5-HTT deficient mice. *PLoS ONE* 6, e22600.
- Pandya, D.N., Van Hoesen, G.W., and Domesick, V.B. (1973). A cingulo-amygdaloid projection in the rhesus monkey. *Brain Res.* 61, 369–373.
- Pannu Hayes, J., Labar, K.S., Petty, C.M., McCarthy, G., and Morey, R.A. (2009). Alterations in the neural circuitry for emotion and attention associated with posttraumatic stress symptomatology. *Psychiatry Res.* 172, 7–15.

- Pape, H.C., and Pare, D. (2010). Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol. Rev.* *90*, 419–463.
- Paré, D., Quirk, G.J., and Ledoux, J.E. (2004). New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* *92*, 1–9.
- Paz, R., Pelletier, J.G., Bauer, E.P., and Paré, D. (2006). Emotional enhancement of memory via amygdala-driven facilitation of rhinal interactions. *Nat. Neurosci.* *9*, 1321–1329.
- Paz, R., Bauer, E.P., and Paré, D. (2007). Learning-related facilitation of rhinal interactions by medial prefrontal inputs. *J. Neurosci.* *27*, 6542–6551.
- Paz, R., Bauer, E.P., and Paré, D. (2009). Measuring correlations and interactions among four simultaneously recorded brain regions during learning. *J. Neurophysiol.* *101*, 2507–2515.
- Peters, J., Dieppa-Perea, L.M., Melendez, L.M., and Quirk, G.J. (2010). Induction of fear extinction with hippocampal-infralimbic BDNF. *Science* *328*, 1288–1290.
- Phelps, E.A., Delgado, M.R., Nearing, K.I., and LeDoux, J.E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* *43*, 897–905.
- Popa, D., Duvarci, S., Popescu, A.T., Léna, C., and Paré, D. (2010). Coherent amygdalocortical theta promotes fear memory consolidation during paradoxical sleep. *Proc. Natl. Acad. Sci. USA* *107*, 6516–6519.
- Rescorla, R.A. (1999). Partial reinforcement reduces the associative change produced by nonreinforcement. *J. Exp. Psychol. Anim. Behav. Process.* *25*, 403–414.
- Roesch, M.R., Calu, D.J., Esber, G.R., and Schoenbaum, G. (2010). Neural correlates of variations in event processing during learning in basolateral amygdala. *J. Neurosci.* *30*, 2464–2471.
- Rushworth, M.F., and Behrens, T.E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* *11*, 389–397.
- Sangha, S., Narayanan, R.T., Bergado-Acosta, J.R., Stork, O., Seidenbecher, T., and Pape, H.C. (2009). Deficiency of the 65 kDa isoform of glutamic acid decarboxylase impairs extinction of cued but not contextual fear memory. *J. Neurosci.* *29*, 15713–15720.
- Shin, L.M., Bush, G., Milad, M.R., Lasko, N.B., Brohawn, K.H., Hughes, K.C., Macklin, M.L., Gold, A.L., Karpf, R.D., Orr, S.P., et al. (2011). Exaggerated activation of dorsal anterior cingulate cortex during cognitive interference: a monozygotic twin study of posttraumatic stress disorder. *Am. J. Psychiatry* *168*, 979–985.
- Sierra-Mercado, D., Padilla-Coreano, N., and Quirk, G.J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology* *36*, 529–538.
- Stefanacci, L., and Amaral, D.G. (2002). Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. *J. Comp. Neurol.* *451*, 301–323.
- Vazdarjanova, A., and McGaugh, J.L. (1998). Basolateral amygdala is not critical for cognitive memory of contextual fear conditioning. *Proc. Natl. Acad. Sci. USA* *95*, 15003–15007.
- Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S.L., and Quirk, G.J. (2006). Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learn. Mem.* *13*, 728–733.
- Wallis, J.D., and Kennerley, S.W. (2010). Heterogeneous reward signals in prefrontal cortex. *Curr. Opin. Neurobiol.* *20*, 191–198.