Neuron
Article

Functional Connectivity between Amygdala and Cingulate Cortex for Adaptive Aversive Learning

Oded Klavir,[1] Rotem Genud-Gabai,[1] and Rony Paz[1,*]
[1]Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel
*Correspondence: rony.paz@weizmann.ac.il
http://dx.doi.org/10.1016/j.neuron.2013.09.035

SUMMARY

The ability to switch flexibly between aversive and neutral behaviors based on predictive cues relies on learning driven by surprise or errors in outcome prediction. Surprise can occur as absolute value of the error (unsigned error) or its direction (signed errors; positive when something unexpected is delivered and negative when something expected is omitted). Signed and unsigned errors coexist in the brain and were associated with different systems, but how they interact and form across large networks remains vague. We recorded simultaneously in the amygdala and dorsal anterior cingulate cortex (dACC) of monkeys performing a reversal aversive-conditioning paradigm and quantified changes in interregional correlations when contingencies shift. We report that errors exist in different magnitudes and that they differentially develop at millisecond resolution. Our results support a model where unsigned errors first develop in the amygdala during successful learning and then propagate into the dACC, where signed errors develop and are distributed back to the amygdala.

INTRODUCTION

Daily life requires behavioral flexibility, and this is crucial when environmental cues become predictive of aversive outcomes and then switch contingencies: an aversive cue becomes neutral, and/or a previously neutral cue becomes predictive of danger. Such associative learning is a flexible process driven by surprise signals—the difference between expected and actual outcome—also known as prediction errors (PEs). Extensive work has suggested two different kinds of PE that can drive learning. In one, the error term poses a directional property, a positive or a negative signed value: positive when something unexpected is delivered and negative when an expected outcome is suddenly omitted (Rescorla and Wagner, 1972; Schultz et al., 1997; Sutton and Barto, 1998). In the other, the effect on associative strength is driven by the absolute value of the error, regardless of its sign, and correlated only with its magnitude (Pearce and Hall, 1980). This kind of error typically reflects an attention signal that can enhance cues' associability (Pearce and Mackintosh, 2010; Roesch et al., 2012; Salzman et al., 2007).

Previous work has pointed to the amygdala as a possible source for absolute errors, potentially through its role in attention (Belova et al., 2007; Li et al., 2011; Murray, 2007; Roesch et al., 2010). Nevertheless, careful observation of the activity of single neurons has revealed that some units respond to omission of aversive outcome and other units respond to omission of appetitive outcome, suggesting that the sign of the error can also be represented (Belova et al., 2007). The dorsal anterior cingulate cortex (dACC) plays a role in the computation of different types of errors (Alexander and Brown, 2011; Kennerley et al., 2011; Matsumoto et al., 2007; Rushworth and Behrens, 2008; Seo and Lee, 2007), as well as absolute error and attention (Bryden et al., 2011; Hayden et al., 2011). Whereas most studies focus on one region at a time, and conclusions are drawn based on different paradigms and from different species, these two regions form a two-way dense-projections network that is extensively evolved in the primate (Ghashghaei et al., 2007; Wise, 2008). Moreover, most studies that examined PEs used an appetitive paradigm (or a mixture of appetitive and aversive), yet the dACC-amygdala network is extensively involved in fear learning and its update (Dunsmoor et al., 2007; Hartley et al., 2011; Klavir et al., 2012; Livneh and Paz, 2012a, 2012b; Milad and Quirk, 2012; Paz and Pare, 2013; Schiller and Delgado, 2010; Sierra-Mercado et al., 2011), and failure to update might underlie anxiety disorders (Etkin et al., 2011; Pitman et al., 2012; Rauch et al., 2006; Shin and Liberzon, 2010). We therefore asked how PEs transfer across this tight network to underlie successful adaptive aversive learning.

To address this, we recorded simultaneously in the amygdala and dACC (dorsal-BA24) of monkeys performing an aversive-conditioning task and its reversal (Li et al., 2011; Schiller et al., 2008). We used the reversal phase to map responses of single units to types of PEs. We then used this to compare the functional connectivity between these regions on a millisecond resolution and how information travels in the network during successful updating of contingencies.

RESULTS

Learning of Aversive Associations and Its Reversal

In each session (n = 96 days, 49 and 47 for monkey B and monkey L, respectively), following habituation to two new stimuli

(a pure tone and a visual fractal, six trials each, interleaved), monkeys associated one stimulus with an air puff to the eye (conditioned stimulus, CS+), and the other stimulus was unpaired (CS−) (15 trials each, interleaved; Figure 1A). Then, a reversal phase switched contingencies between the CS+ and the CS− (15 trials each). Learning was quantified as the conditioned response (CR), a preparatory eye blink measured by the total time the eye was closed during the 400 ms following CS offset until 100 ms prior to US delivery (Figure 1B). This revealed no difference during habituation (Figure 1C; CS main effect, $F[1, 192] = 0.29$, $p = 0.6$) but that monkeys learned to differentiate by responding more to the CS+ during acquisition (CS main effect, $F[1, 194] = 20.14$, $p < 0.001$) and to reverse contingencies and respond more to the other CS during reversal (CS main effect, $F[1, 190] = 4.94$, $p < 0.05$). We made sure that reversal was surprising on a daily basis by comparing early versus late sessions (notice also that learning curves were gradual during reversal rather than an immediate switch, Figure 1C), and by including control sessions with more CSs−, so that they could not predict which is the new CS+.

Amygdala and dACC Responses Represent PEs

We recorded 95 CS-responsive dACC neurons (Figures 1D and 1F; 50 and 45 for monkeys B and L, respectively) and 122 amygdala neurons (Figures 1E and 1G; 67 and 55 for monkeys B and L, respectively). Neurons were considered CS responsive if they had a significant change in response to either CS at either phase by comparing firing rates (FRs) in the 1,000 ms after CS offset to the baseline taken from 1,000 ms before CS onset (paired t test over trials, $p < 0.05$). Each neuron FR was then normalized to its own overall average FR for all further analyses. Cells were mainly recorded in the B/AB (BL/BM) nuclei of the amygdala and in dorsal-BA24 of the dACC (Figures 1F and 1G), as these regions are extensively and bidirectionally connected in the primate (Ghashghaei et al., 2007).

Each neuron was classified into a PE type. To do so, we calculated two indices for each neuron: one index was based on the neural response to the CS+ that turned into a CS− (negative index), and the other index was for the response to the CS− that turned into a CS+ (positive index). These indices measure the change in activity for the CS from late learning (average FR from last three trials) to early reversal (average FR from first three trials). An index was calculated as (reversal – acquisition)/(reversal + acquisition) and, hence, ranges from −1 to 1, with values of >0, indicating that the neuron fired more during early reversal than during late acquisition, and vice versa for values <0. Therefore, the negative index quantifies negative error (NEr), because it measures changes from predicting an expected puff (late acquisition) to its omission (early reversal), and the positive index measures changes from no outcome to a surprising puff, a positive error (PEr). If a neuron responds similarly to both PEr and NEr (>0 in both indices), it signals unsigned PE (an absolute value); otherwise, it signals signed PE, being either NEr or PEr oriented (Figure S1 available online).

We found that, in both regions, units signaled more of a change from no puff to puff: amygdala, mean of index, 0.18, $t(121) = 4.25$, $p < 0.001$ (Figure 2A, upper histogram); dACC, mean, 0.11, $t(94) = 2.31$, $p < 0.05$, t tests (Figure 2B). For a change from puff to no puff, there was homogenous representation in the dACC: mean, −0.07, $t(94) = -0.94$, $p > 0.1$ (Figure 2B, right histogram); but there was less representation in the amygdala: mean, −0.13, $t(121) = -3.12$, $p < 0.01$ (Figure 2A). Similar results are obtained when comparing the actual number of neurons (Figures 2A and 2B, number and percentage of neurons indicated next to the axes, binomial tests, $p < 0.01$). These proportions of units remained similar across amygdala and dACC when we used late versus early reversal for computing the indices (Figure S2).

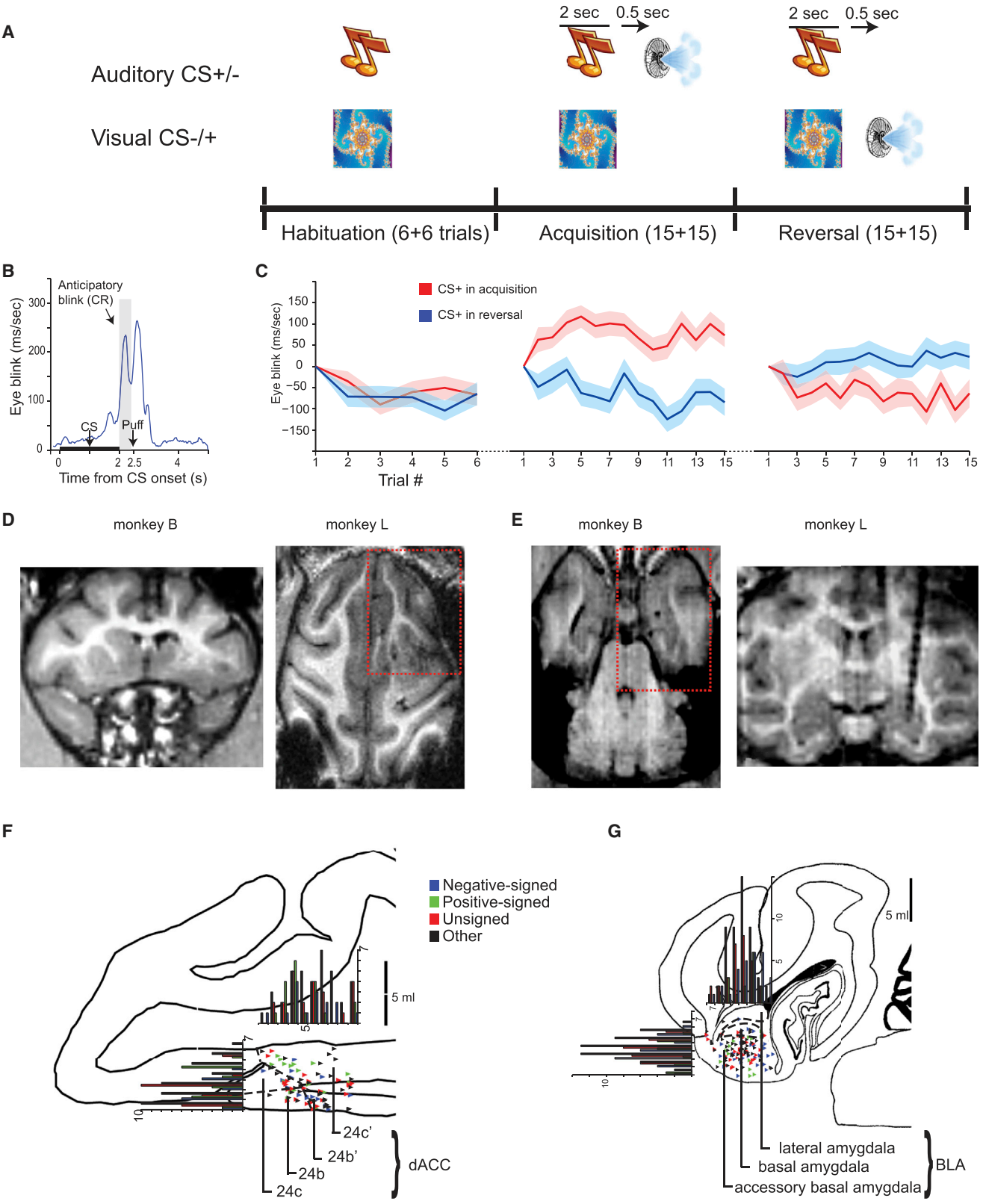
To make sure these changes reflect surprise signals and not a response to air puff or its absence, we classified a neuron as PEr only if it also increased its response to the CS+ during early reversal when compared to the CS+ during late acquisition (a puff index; see Experimental Procedures and different shapes in Figures 2A and 2B; hence, a portion of its increased response can be attributed to surprise) and complementary for NEr− only if it responded more to the CS− during early reversal than to the CS− in late acquisition (a no-puff index; hence, the response can be attributed also to surprise and not only to puff absence).

In the amygdala, 29 units were classified as unsigned error and 31 as signed; of the signed ones, only 3 were NEr and 28 were PEr (Figures 2A and 2C). In the dACC, 23 were unsigned and 26 were signed: 11 NEr and 15 PEr (Figures 2B and 2D). There was no significant difference in the proportion of either error type across the two regions (Figure 2E; unsigned: 23.8%/24.2% for amygdala/dACC, $p = 0.9$; signed: 25.4%/27.4%, $p = 0.9$, chi-square).

Furthermore, differences can occur in the magnitude of the overall signal, rather than in classification of neurons. We therefore quantified the overall percent change in FR. We found, first, that amygdala neurons exhibited a larger unsigned signal (Figure 2F, left) (amygdala, 133.53%; dACC, 98.52%); second, that FR changes were larger for NEr in the dACC (Figure 2F, middle) (amygdala, 14.52%; dACC, 103.11%); and finally, that FR changes were larger for PEr in the amygdala (Figure 2F, right) (amygdala, 261.77%; dACC, 50.58%, $t[41] = 2.53$, $p = 0.02$). Notice that PErs quantify preference for negative valence in this aversive set-up (a surprise addition of air puff), potentially explaining the stronger amygdala response. This is also in line with the much higher number of neurons in the amygdala that represent PEr than NEr (28 versus 3, $p < 0.01$, chi-square). We conclude that, although the amygdala represents signed errors in addition to unsigned (absolute) errors, it is biased toward representation of the surprise addition of aversive information.

Changes in Amygdala-dACC Connectivity Dissociate Successful Learning

To examine functional connectivity and how it changes during successful learning, we measured cross-correlations (CCs; Figure 3A) between all simultaneously recorded pairs of neurons from both regions. We were specifically interested here in the directionality of information transfer and, therefore, calculated the distribution of center of mass (CoM) for all significant CCs ($n = 229$ during habituation, and $n = 302$ during learning, significance tested against shuffled CC). We always used the amygdala unit as a reference (time zero) when calculating CCs to allow



(legend on next page)

easy interpretation and presentation; hence, a positive CoM means that amygdala activity precedes, and a negative CoM means that dACC precedes. We found that the overall baseline distribution of CoMs during the habituation phase was significantly shifted to indicate that dACC activity precedes amygdala but became less so during learning (Figure 3B): habituation, −3.71 ± 1.95 ms, t(222) = −3.75, $p < 0.001$; learning, −0.80 ± 1.06 ms, t(297) = −1.49, $p = 0.14$. In addition to directionality, we found a major decrease in the variance of the population of CoM, indicating a more precise locking and functional interactions that emerge during learning (Figure 3B, inset), F(222, 297) = 2.50, $p < 0.0001$. Notice that the common CoM is of few milliseconds, which is indicative of this tight one-synapse network (Ghashghaei et al., 2007).

The focus of this study was to test how directionality of information transfer in this reciprocal network changes as a function of three factors: (1) learning (whether it was successful or not: to do this, we separated sessions in which discrimination was learned significantly during reversal [n = 51 days with CR significantly different across CS+ and CS−, $p < 0.05$, t test] from sessions in which it was not learned significantly [n = 49 nondiscriminating days, $p > 0.05$]); (2) error type (signed or absolute errors, according to the classification in the previous section); and (3) source of the PE (amygdala/dACC: grouping to PE types based on the classification of the amygdala neuron in the pair or the dACC neuron [notice that these are not mutually exclusive groups]). We found that the full model that compared CoM across the three factors was significant (three-way ANOVA, F [1,175] = 10.8, $p < 0.01$) and further revealed several specific functional findings.

First, amygdala unsigned units precede the dACC, but only during successful learning when compared to nondiscriminating days (3.37 ms versus −4.39 ms, $p < 0.01$, least significant difference [LSD] post hoc tests; Figure 3C, dark blue versus light blue). Second, during successful learning, amygdala unsigned units precede when compared to signed units—where the dACC precedes (3.37 ms versus −1.45 ms, $p < 0.05$, LSD; Figure 3C, dark blue versus dark red).

Therefore, successful learning is characterized by unsigned amygdala units that precede dACC activity and dACC activity that precedes signed amygdala units. Complementing this, we further found that signed dACC units are preceded by amygdala activity during successful learning when compared to nondiscriminating sessions (1.48 ms versus −3.7 ms, $p = 0.05$, LSD; Figure 3D, dark red versus light red). These shifts in CoM for the different conditions are further clarified when normalizing to the baseline delay during habituation (Figure S3).

To summarize these effects, we created contrasts between discriminating (successful learning) and nondiscriminating days in all conditions. This revealed the double dissociation where amygdala precedes for unsigned amygdala units and for signed dACC units (Figure 4A, 7.75 ms and 5.18 ms, leftmost and rightmost bars, $p < 0.01$, F test). Finally, we contrasted error types only during successful learning sessions, revealing dissociation between amygdala and dACC for unsigned minus signed units (Figure 4B; amygdala led by 4.82 ms and dACC led by 4.19 ms).

Overall, these findings suggest that, during successful learning only, amygdala unsigned units precede, and might be required for, formation of signed dACC units, and then signed amygdala units are driven by these dACC inputs.

DISCUSSION

In this study, we recorded simultaneously the responses of units in the primate amygdala and the dACC during reversal of aversive conditioning. This allowed a direct comparison of how information about error types travels in this network at a millisecond resolution and how changes in functional connectivity underlie successful learning. We found that both types of surprise signals (absolute and signed) are common in both regions, with differences in the magnitude of the signal. Using functional correlations, we found that activity in dACC leads in baseline conditions and in days when learning was weak. During successful learning, however, unsigned error in the amygdala precedes dACC activity, whereas dACC activity precedes signed errors in the amygdala. Summarizing the findings, our results suggest a model in which successful discriminative aversive learning requires changes in functional connectivity where attentional signals first occur in the amygdala and then propagate into the dACC where signed errors occur; these, in turn, propagate back into the amygdala.

These findings are relevant from two converging aspects discussed later. The first involves flexibility and adaptive aversive learning, and the second involves signals of learning in general.

Figure 1. Behavioral Paradigm and Recording Locations

(A) Each session included two new stimuli from different modalities, one auditory and one visual. Each day started with a habituation phase when stimuli are presented without outcome (six trials each, randomly interleaved). This was followed by an acquisition phase when one of the stimuli, the CS+, was followed by an air puff to the eye in a trace-conditioning design. A reversal phase followed in which the CS+ now turned into CS− and the CS− became the CS+ and was followed by an air puff.

(B) The behavioral learned CR, averaged over all sessions in the trace-conditioning paradigm. The CS was presented for 2 s, and the air puff was delivered 500 ms afterward. The animals started to blink toward the end of the CS and during the trace interval.

(C) CR to the two CSs during habituation, acquisition, and reversal averaged over all sessions (±SEM in shaded colors). While there is no difference during habituation, in both acquisition and reversal there is an increased differential response to the CS+. See Results for statistics.

(D and E) Anatomical MRI scans of transverse and coronal sections through the dACC (D) and the amygdala (E) for the two animals. MRI scans were performed before, during, and after the recording period with calibrating electrodes that reach either the dACC or the amygdala (seen in the figures as dark shades) and were used for alignment of daily penetrations. The red dotted frames are the areas sketched in the schemes in (F) and (G).

(F and G) Scheme of anatomical sections of the macaque brain (transverse section), with recording sites reconstructed based on MRI with calibrating electrodes. Dots and histograms show anatomical locations of responding neurons in the different error groups (for definition of the error groups, see Figure 2). Locations were distributed homogenously in the dACC (F) and amygdala (G) both in the mediolateral plane and in the anterior-posterior plane (chi-square test, all ps > 0.05). BLA, basolateral complex of the amygdala.

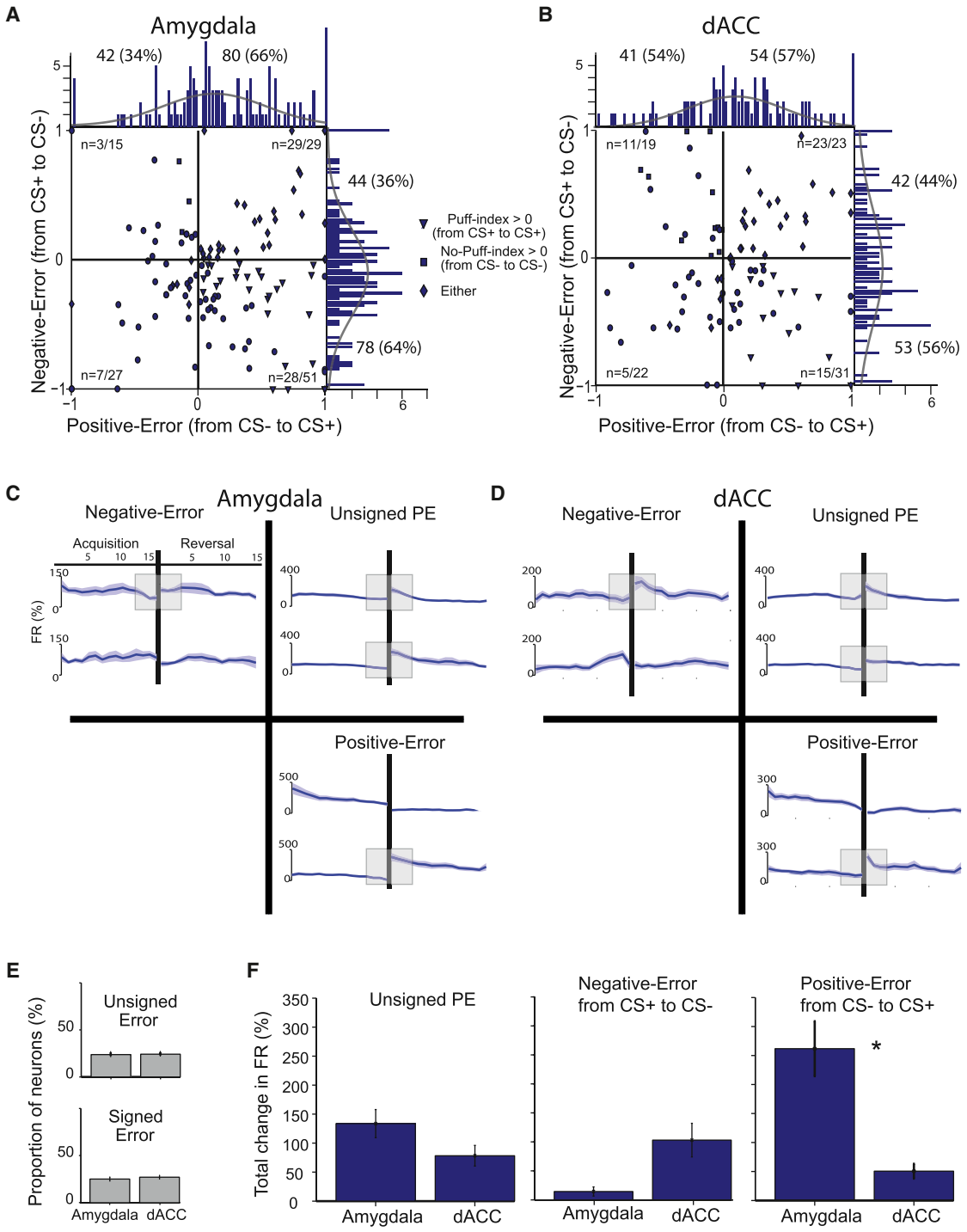


Figure 2. Distribution of PEs in the Amygdala and dACC

(A and B) Amygdala (A) and dACC (B) neurons were classified into PE types based on two indices: positive index (x axis), and negative index (y axis), calculated from comparing activity at the end of acquisition to activity during early reversal, when the two CSs change contingencies. Marginal distributions for each index with fitted Gaussians are shown, and proportions of neurons are mentioned next to them. Additional constraints made sure that these were surprise signals and not only responses to a puff or its absence, comparing CS+ to CS+ and/or CS− to CS− during early reversal versus late acquisition ([no]-puff indices; see Experimental Procedures for details). Triangles represent units that responded more to the CS+ during reversal than to the CS+ during acquisition (puff index > 0); hence, their response to the CS− that turned into CS+ (positive index) is not due only to the puff. Squares represent the same when comparing CS− to CS− (no-puff index > 0); hence the no-puff index restricts the negative index for surprise signals and not only preference for absence of puff. Diamonds indicate either a puff index > 0 or a no-puff index > 0. Circles are neither (all the rest). The conjunctions result in classification to unsigned errors (diamonds in upper right quartile),

(legend continued on next page)

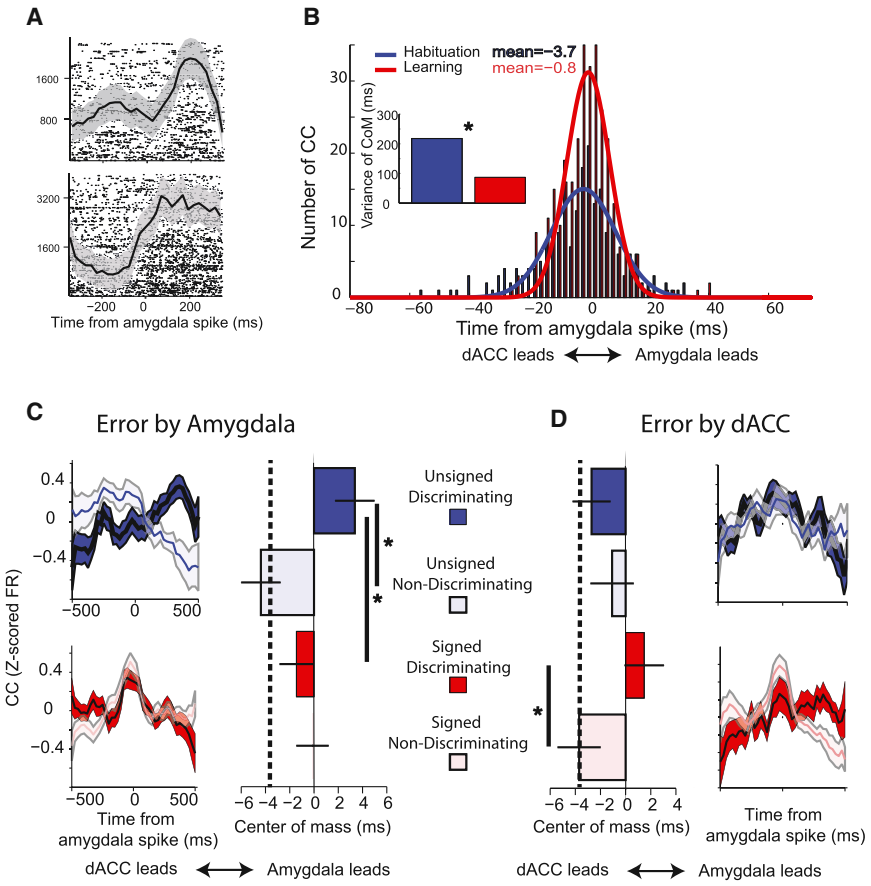


Figure 3. Directionality of CCs Reveals Functional Connectivity

(A) Two examples of raw CCs between simultaneously recorded amygdala and dACC units. Amygdala spikes are at time 0, and raster plots show dACC spikes.

(B) Histogram of the CoMs from all significant cross-regional CCs. All CCs are calculated so that amygdala spike is at time 0; hence, a positive CoM indicates that the amygdala unit leads, and a negative CoM indicates that the dACC unit leads. CCs were calculated separately during habituation (blue) and during learning (red). In both cases, the dACC activity slightly preceded that of the amygdala, but during learning, the variance of the distributions was significantly reduced (inset), suggesting more precise interactions and locking during learning.

(C and D) The CoM for all CCs was calculated separately for three categories: successful learning versus nondiscriminating days (dark versus light colors); signed versus unsigned PE (red versus blue); and when the pair is classified for PE according to the amygdala unit (C) or the dACC unit (D). For each category, shown is the overall averaged CC (with ±SEM shaded) and the corresponding mean of all CoMs in the bars (±SEM). The full three-way model was significant. Asterisks indicate significant post hoc comparisons. The dotted line indicates the mean directionality (CoM) during habituation, as in (B), where dACC leads by 3.7 ms.

These aspects are tightly linked in our study. Most evidence about error types comes from studies involving appetitive paradigms (reward) (Roesch et al., 2012; Schultz, 2012) or a mixture of aversive and appetitive outcomes (Belova et al., 2007; Joshua et al., 2009; Matsumoto and Hikosaka, 2009; Paton et al., 2006). Less is known about characterization of error types in single neurons in a purely aversive paradigm.

Concerning the first aspect, the amygdala-prefrontal network is implicated in fear conditioning and its expression, and is crucial for the updating of associations. Mainly, this network plays an important role in expression of aversive memories and their extinction (Dunsmoor et al., 2007; Herry et al., 2010; Klavir et al., 2012; Linnman et al., 2012; Pare and Duvarci, 2012; Pitman et al., 2012; Schultz et al., 2012; Sotres-Bayon and Quirk, 2010), and also in reversal (Chudasama et al., 2012; Morris and Dolan, 2004; Schiller et al., 2008) and cognitive modulation (Ochsner and Gross, 2005; Schiller and Delgado, 2010). Moreover, failure to update such memories is linked to abnormal functionality and integrity of the dACC (Milad et al., 2009; Shin et al., 2011), likely through the effect it exerts on the amygdala and other output stations. Hence, the ability to update an aversive memory likely relies on the processing of learning signals in this network and might be linked to the success of behavioral therapy for anxiety disorders (Lee, 2013; Pitman et al., 2012; Salzman and Fusi, 2010).

Much research on adaptive aversive learning focused on extinction of fear memories (Milad and Quirk, 2012). The wealth

signed PErs (triangles in lower right quartile), and signed NErs (squares in upper left quartile). The actual number of classified neurons is marked within each quartile along with the total number of units.

(C and D) FR changes from baseline (shaded blue represent ±SEM) for units of different error types in the two regions: amygdala in (C) and dACC in (D). Each type of PE appears on the quadrant matching its location on the coordinate system corresponding to the PE indices, as in (A) and (B). In each quadrant shown is the change in FR to one CS throughout the whole acquisition and reversal periods (15 trials each). The upper row in each quadrant is for the CS+ that turns into CS− during reversal, and the lower is for the CS− that turns into a CS+. Change in activity for both is required to classify to PE type; see (A) and (B), Results, and Experimental Procedures). The highlighted gray square represents the period taken to classify the change in activity (last three trials of acquisition and first three trials of reversal; only significant changes are highlighted). For example, for unsigned/absolute errors, shown in the upper right quartiles, notice that the increase in early reversal is evident for both CSs (upper and lower rows) and is slightly larger in effect magnitude in the amygdala, as shown in (C). See Results for statistics.

(E) Percentage of unsigned units in the amygdala and dACC (upper bars) and signed units (lower bars). There was no significant difference between the regions.

(F) Comparing the magnitude of the change in FR. Shown is the percent change from baseline, in early reversal minus late acquisition, averaged over all relevant units (±SEM). Amygdala unsigned units responded more vigorously (left), as well as amygdala PErs (right), and dACC NErs (middle).

See also Figure S1.

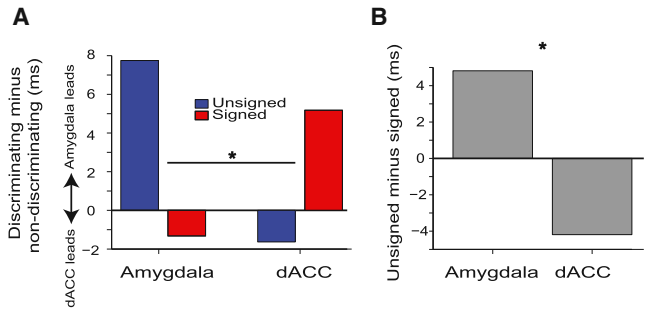


Figure 4. Contrasts Summarizing the Shifts in CoM for the Different Conditions

(A) The difference between discriminating (successful learning) and non-discriminating days, for both PE types and both regions. A double dissociation across regions and PE types is observed ($p < 0.01$, F test).

(B) The difference in CoM between unsigned and signed PEs during successful learning only.

of evidence shows that the medial-prefrontal-cortex (mPFC) can be grossly divided into two zones (Sierra-Mercado et al., 2011): one inhibiting the amygdala during fear extinction—the rat infralimbic cortex (IL), paralleled by the human ventromedial-prefrontal-cortex (vmPFC); and the other supporting fear expression and its learning—the rat prelimbic cortex (PL), which is most likely paralleled by the primate dACC and the region we recorded from here. Here, we used a full reversal paradigm that is required to identify PEs, as extinction alone is not enough because a CS′ value changes only in one direction: from positive to negative. Reversal requires creating a new aversive association in parallel to extinguishing an existing one. There are two reasons to hypothesize that the anterior cingulate cortex (ACC) would be involved in this paradigm: first, because it plays a role in fear acquisition and expression (as for the CS− that turns into a CS+); and second, because it was shown to be involved in complex learning and understanding of the task structure, as well as signaling different error types during learning. Therefore, the more complex learning of switching contingencies that occurs during reversal can be processed in the ACC. Indeed, we found that errors in the ACC and its connectivity with the amygdala support successful reversal. However, this opens an interesting question: is there a difference in underlying mechanisms between a CS+ that turns into a CS− during extinction, from the same process during reversal? There are two main alternatives: the first possibility is, just like extinction, reversal involves new inhibitory learning for the CS+ that turns into CS−. This would suggest similar modulation by the rodent IL/primate vmPFC; and in parallel, the PL/ACC and the amygdala together contribute to learning of the CS− that turns into a CS+. The second possibility is that the value is transferred between the two CSs during reversal, without direct inhibition of the amygdala, and the ACC takes a more active role by signaling both types of errors and communicating them into the amygdala. These two options are not mutually exclusive.

There is scarce behavioral evidence showing spontaneous recovery and renewal of the original associations after reversal (Bouton, 1993; Rescorla, 2007). This might suggest that inhibition of the original CS+ memory indeed occurs, and additional evidence comes from partial similarities in brain activations between extinction and reversal (Schiller and Delgado, 2010; Schiller et al., 2008). Our paradigm and results cannot dissociate the two aforementioned options, but they do show that ACC-amygdala connectivity is important for the successful learning of reversal. This might, at face value, support the idea that this network underlies complex learning (Livneh and Paz, 2012a), rather than the independent extinction plus new acquisition possibility. However, a direct study comparing extinction and reversal within the same paradigm is required to determine differences and similarities in network mechanisms at the single-cell level. Moreover, it would be interesting to see if reversal (or other complex learning paradigms) might engage this network more efficiently and improve successful updating of the aversive memory in psychopathological conditions.

The second major aspect of this study concerns learning from surprising outcomes more generally (Bouton, 2006; Kamin, 1969), formally defined as PEs. Different models suggest different kinds of teachers. In one class, it is the absolute value of the PE that modulates learning (unsigned error) (Pearce and Hall, 1980): by modulating the amount of attention dedicated to the cues, it modulates their associative strength (associability) (Pearce and Mackintosh, 2010). In another class of models, it is the direction of the error that drives learning (signed error): positive when something unexpected is delivered and negative when something expected is omitted (Rescorla and Wagner, 1972; Schultz et al., 1997; Sutton and Barto, 1998). Although separate models, recent evidence and theory suggest that the two models coexist in the brain and possibly interact to modulate one another (Esber and Haselgrove, 2011; Le Pelley, 2004; Pearce and Mackintosh, 2010; Roesch et al., 2012). Our results show that the two types of errors, signed and unsigned, indeed coexist in the brain; moreover, they coexist within specific regions—the amygdala and the ACC. On one hand, this could be evidence that the two models coexist or even compete in parallel in local networks. On the other hand, we show here that the two signals are coupled within millisecond resolution between these two regions, with specific functional directionality: unsigned amygdala precedes dACC and dACC precedes signed amygdala. These findings can support an integrated model in which shifts in attention signals coming from the amygdala (Belova et al., 2007; Johansen et al., 2010; Murray, 2007; Roesch et al., 2010; Tye et al., 2010), are used to scale the signed errors that develop in the dACC (Behrens et al., 2007; Courville et al., 2006; Kakade and Dayan, 2002; Preuschoff and Bossaerts, 2007). Our results further agree with the proposal that error signals in the dACC can originate from amygdala inputs (Belova et al., 2008; Roesch et al., 2012), and this applies for attention signals in the dACC (Bryden et al., 2011; Hayden et al., 2011) as well as for other types of errors found in prefrontal regions (Alexander and Brown, 2011; Daw et al., 2005; Matsumoto et al., 2007; Paus, 2001; Rushworth and Behrens, 2008; Seo and Lee, 2007; Wallis and Kennerley, 2011).

One noted difference between the current findings and recent reports is the existence of widespread signed errors in the amygdala. Single-cell studies in rodents provided evidence that unsigned attention signals are widely represented in the amygdala, with little or fewer signed errors (Roesch et al.,

2010). We suggest that the difference might stem from our use of a purely aversive task, compared to an appetitive paradigm. Such aversive learning might engage the amygdala to a greater extent (Davis and Whalen, 2001), via tight ACC → amygdala projections as found here (that dACC activity precedes signed amygdala units). In other words, it might be that, during aversive learning, signed errors are communicated into the amygdala more than during appetitive paradigms. In addition, a recent imaging study found associability signal (driven by cue-specific attention, hence, representing more unsigned errors) in the amygdala (Li et al., 2011). Because we found that attention signals appear first in the amygdala during successful learning, and with higher overall activity, these factors could influence the BOLD signal and reconcile the results. Finally, we classified neurons into PE groups independently of the response magnitude; i.e., a neuron would be classified similarly if it has a strong response or a mild response, as long as it was in the same direction. This could create a bias if signed errors in the amygdala have weaker overall activity changes, as we indeed found. An analysis that would take into account response strength before classifying a neuron might also explain the differential findings. Here, however, we wanted to examine the functional interactions between error types and, therefore, included all classified neurons that can contribute to the overall error type signal originating from a region.

The main strength of this study lies in the ability to measure how these two error signals propagate in the network at a millisecond resolution, by calculating CoM of interregional pairwise CCs. The finding that the dACC precedes during baseline conditions (habituation) is in accordance with the heavier projections from dACC to amygdala than vice versa (Ghashghaei et al., 2007) and also parallels recent functional studies (E. Likhtik et al., 2011, Soc. Neurosci., conference; Livneh et al., 2012). The fact that dACC continued to precede during learning when it was weak/unsuccessful is further in line with the role of the dACC in general anxiety (Etkin et al., 2011; Pitman et al., 2012) and can support two leading models of anxiety disorders: a failure to update value from danger to safety, as in extinction (Milad and Quirk, 2012), and overgeneralization of fear to similar stimuli, evidenced by weak nondiscriminatory learning here (Dunsmoor and LaBar, 2013; Laufer and Paz, 2012; Lissek, 2012; Resnik et al., 2011).

Overall, our study shows that successful learning and updating of aversive memories requires both types of error signals, with unsigned errors in the amygdala driving dACC activity and vice versa for signed errors. Better understanding of the exact balance between these signals in different behavioral paradigms and over more brain regions would provide a better picture of how and why the updating of information can fail and lead to maladaptive behaviors.

EXPERIMENTAL PROCEDURES

Animals

Two male *Macaca fascicularis* (4–7 kg) were implanted with a recording chamber (27 × 27 mm) above the right ACC (Broadman 24/32) and 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, following National Institutes of Health regulations and with accreditation from the Association for Assessment and Accreditation of Laboratory Animal Care International. 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 session, the monkeys engaged in a classical trace conditioning task with a random intertrial interval (ITI) of 25 s on average (Figure 1A). The CS+ of either a new pure tone or a new fractal cue (Chaos Pro 4.0 program; http://www.chaospro.de) was paired with an aversive US of an air puff (150 ms duration; three to five bars; located proximally 5 cm from the left eye). Each CS was presented for 2 s, followed by a trace period of 0.5 s and then delivery of the US. The stimulus of the other modality was presented without being followed by any outcome (CS−). Hence, in each session, there was an auditory or a visual CS+ (in randomly interleaved days) and a CS− of the other modality.

Each new session started with a habituation stage, in which the two CSs were presented randomly (six trials each), followed by the acquisition stage described earlier (15 trials per each CS, randomly interleaved) and a reversal stage, in which the CSs switched value, with the CS− becoming a CS+ (followed by the US) and the CS+ becoming a CS− (presented without US) (15 trials each).

There were 96 sessions overall that included full acquisition and reversal, 49 in monkey B and 47 in monkey L.

Behavior

CR was quantified as the anticipatory eye blinks, measured as the total time the eye was closed during the 400 ms starting from CS offset to 100 ms prior to US delivery (Figure 1B).

A digital video camera for dark (infrared) conditions (Provision-ISR) recorded the monkey's left eye at 50 Hz. Video analyses was performed semiautomatically using custom-made software implemented in Matlab to identify periods when the animal closed the eye. In short, the minimal rectangle surrounding the eye was defined by the experimenter for each session, and few typical "closed" states and few typical "open" states were identified. The software then quantified the distribution of pixels' light intensity for each state and then calculated the Jensen-Shannon divergence for each frame, resulting in a probability of being in a closed state and a probability of being in an open state. A threshold for both probabilities resulted in a per-frame classification. We validated the algorithm by random samples of days and found it to be consistent with the judgments of a blind human observer for >95% of the reported eye blinks.

Each day was sorted using a Wilcoxon rank-sum test, comparing the CR from the two CSs during reversal. Discriminating days were defined as days in which the monkey significantly ($p < 0.05$) differentiated between the two CSs in last seven trials (per CS) of reversal training. Nondiscriminating days were defined as all other days.

It could be argued that monkeys could learn that reversal occurs daily and predict it. However, it is probably impossible for the animal to estimate when exactly this would happen, because there were 30 trials during acquisition with variable ITI, and we doubt that the animal can count it. To assure this, we separated early sessions from late sessions and tested whether behavior was similar at the end of acquisition and at early reversal. There was no difference in differentiation of CSs at early reversal in early sessions versus late sessions ($p > 0.1$, t test), indicating that the reversal came as a surprise. In addition, we repeated the paradigm in sessions in which two additional CSs− (one of each modality) were used (Klavir et al., 2012); hence, the animal could not know which CS− would reverse to be a CS+. There was no difference in learning behavior in these days. Notice also the gradual learning curves in Figure 1, showing that real learning occurred. In summary, although we repeated the paradigm for several sessions in each animal, early reversal was a surprise, as also reflected in the neural activity. Finally, we used completely new stimuli each day (different fractals and different pure tones).

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). A T1-weighted, three-dimensional gradient-echo (MPRAGE) pulse sequence was acquired with a repetition time of 2,500 ms, an inversion time of 1,100 ms, an echo time of 3.36 ms, an 8° flip angle, and two averages. Images were acquired in the sagittal plane, 192 × 192 matrix, and 0.6^3 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 dACC, amygdala, and anatomical markers such as the interaural line and the anterior commissure; confirmed using atlas (Martin and Bowden, 2000; 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 dACC and the amygdala, and two to three observers separately inspected the images and calculated the regions' anterior-posterior and lateral-medial borders relative to the electrodes. The depth of the two regions was calculated from the dura surface.

Recordings

The monkeys were seated in a dark room and each day, up 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 once in the cortex. The electrodes were then moved independently further into either the dACC or the amygdala (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 dACC and amygdala). Electrode signals were preamplified, 0.3 Hz–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 further performed for all sessions to improve unit isolation (offline sorter, Plexon).

Data Analyses

Response FR was computed as the average FR in the 1,000 ms from CS offset. The response therefore includes all of the Trace interval and the immediate response to US delivery, as in previous relevant studies (Belova et al., 2007; Calu et al., 2010; Roesch et al., 2010). Responsive units were defined by comparing this FR to the baseline taken 1,000 ms prior to CS onset, using a paired t test. The response FR was then normalized to the average FR during the learning phase and used for all subsequent analyses.

PE Groups

To sort the units into PE groups, the average FR change was calculated for each CS in the last three trials of acquisition and in the first three trials of reversal. Error indices were calculated as follows:

$$\frac{FR(early\_reversal) - FR(late\_acquisition)}{FR(early\_reversal) + FR(late\_acquisition)},$$

providing a measure ranging from −1 to 1. This index was calculated twice, once for each CS. The NEr index was calculated for the CS that turned from positive to negative; i.e., the air puff previously following the CS is now omitted for the same CS, hence, a NEr. The PEr index was defined for the CS that turned from negative to positive; i.e., the previously neutral "safe" CS is now followed by an air puff, hence, a PEr.

Units were then classified as follows:

(1) Unsigned PE: positive index > 0 and negative index > 0.
(2) Signed PE:
- NEr: negative index > 0 and positive index < 0.
- PEr: negative index < 0 and positive index > 0.

However, note that some neurons could simply be responding more to the addition of an air puff independent of surprise, or even to its absence independent of surprise. Hence, we applied an additional constraint and calculated a puff index, defined by the response to CS+ in early reversal compared to the response to CS+ in late acquisition (i.e., the two different CSs but when both are followed by an air puff) and a no-puff index similarly defined on the response to CS− in early reversal compared to the response to CS− in late acquisition. For the PEr, we required a puff index > 0 to assure that it is not merely a response to the air puff. In other words, to have a PEr, it is not enough that a neuron responds more to the CS+ compared to when it was a CS−, but it also has to respond to the puff more than before; therefore, at least some component of its increased response is due to the surprise. For NEr, we required that a no-puff index > 0, hence some of the neuron response is related to surprise and not only to the absence of a puff. For unsigned error, we required that either index (puff or no puff) be > 0; hence, some component of the neural response is due to surprise for a CS.

The remaining units were not assigned. It could be argued that units with a positive index and negative index < 0 are also unsigned inhibition. We made sure that it did not change the major findings in any way and preferred to avoid treating them here.

CCs

For each pair of amygdala-dACC units that were recorded simultaneously (i.e., during the same session), standard CCs were calculated in windows of −500 to 500 ms around each amygdala spike and in 25 ms bins. Shuffling technique was used to assess statistical significance: we shuffled the order of the trials 40 times and computed CC at the shuffled condition. CCs that included clusters of bins that exceeded 95% of the shuffled distribution were identified as significant.

For each CC, a CoM was computed as follows:

$$\frac{\sum^{(FR_i * T_i)}}{\sum^{(FR_i)}}.$$

All results concerning behavior were analyzed using a repeated-measures ANOVA with eye blink width as dependent variable, Significant interactions were followed by post hoc LSD comparisons, and for all comparisons, significance was assumed at $p < 0.05$, unless reported specifically.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2013.09.035.

ACKNOWLEDGMENTS

We thank Yossi Shohat for his valuable contributions 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. The work was supported by the I-CORE Program of the Israel Science Foundation (grant 51/11) and by ERC-FP7-StG 281171 grant to R.P.

Accepted: September 9, 2013
Published: December 4, 2013

REFERENCES

Alexander, W.H., and Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. Nat. Neurosci. *14*, 1338–1344.

Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

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.

Belova, M.A., Paton, J.J., and Salzman, C.D. (2008). Moment-to-moment tracking of state value in the amygdala. J. Neurosci. *28*, 10023–10030.

Bouton, M.E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. Psychol. Bull. *114*, 80–99.

Bouton, M.E. (2006). Learning and Behavior: A Contemporary Synthesis. (Sunderland: Sinauer Associates).

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.

Calu, D.J., Roesch, M.R., Haney, R.Z., Holland, P.C., and Schoenbaum, G. (2010). Neural correlates of variations in event processing during learning in central nucleus of amygdala. Neuron *68*, 991–1001.

Chudasama, Y., Daniels, T.E., Gorrin, D.P., Rhodes, S.E., Rudebeck, P.H., and Murray, E.A. (2012). The Role of the Anterior Cingulate Cortex in Choices based on Reward Value and Reward Contingency. Cereb. Cortex. Published online September 3, 2012. http://dx.doi.org/10.1093/cercor/bhs266.

Courville, A.C., Daw, N.D., and Touretzky, D.S. (2006). Bayesian theories of conditioning in a changing world. Trends Cogn. Sci. *10*, 294–300.

Davis, M., and Whalen, P.J. (2001). The amygdala: vigilance and emotion. Mol. Psychiatry *6*, 13–34.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

Dunsmoor, J.E., and LaBar, K.S. (2013). Effects of discrimination training on fear generalization gradients and perceptual classification in humans. Behav. Neurosci. *127*, 350–356.

Dunsmoor, J.E., Bandettini, P.A., and Knight, D.C. (2007). Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. Behav. Neurosci. *121*, 635–642.

Esber, G.R., and Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. Proc. Biol. Sci. *278*, 2553–2561.

Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci. *15*, 85–93.

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.

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.

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., Ferraguti, F., Singewald, N., Letzkus, J.J., Ehrlich, I., and Lüthi, A. (2010). Neuronal circuits of fear extinction. Eur. J. Neurosci. *31*, 599–612.

Johansen, J.P., Tarpley, J.W., LeDoux, J.E., and Blair, H.T. (2010). Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. Nat. Neurosci. *13*, 979–986.

Joshua, M., Adler, A., Rosin, B., Vaadia, E., and Bergman, H. (2009). Encoding of probabilistic rewarding and aversive events by pallidal and nigral neurons. J. Neurophysiol. *101*, 758–772.

Kakade, S., and Dayan, P. (2002). Acquisition and extinction in autoshaping. Psychol. Rev. *109*, 533–544.

Kamin, L.J. (1969). Predictability, surprise, attention and conditioning. In Punishment and Aversive Behavior, B.A. Campbell and R.M. Church, eds. (New York: Appleton-Century-Crofts), pp. 279–296.

Kennerley, S.W., Behrens, T.E.J., and Wallis, J.D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat. Neurosci. *14*, 1581–1589.

Klavir, O., Genud-Gabai, R., and Paz, R. (2012). Low-frequency stimulation depresses the primate anterior-cingulate-cortex and prevents spontaneous recovery of aversive memories. J. Neurosci. *32*, 8589–8597.

Laufer, O., and Paz, R. (2012). Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. J. Neurosci. *32*, 6304–6311.

Lee, D. (2013). Decision making: from neuroscience to psychiatry. Neuron *78*, 233–248.

Le Pelley, M.E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. Q. J. Exp. Psychol. B *57*, 193–243.

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.

Linnman, C., Zeidan, M.A., Furtak, S.C., Pitman, R.K., Quirk, G.J., and Milad, M.R. (2012). Resting amygdala and medial prefrontal metabolism predicts functional activation of the fear extinction circuit. Am. J. Psychiatry *169*, 415–423.

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.

Livneh, U., and Paz, R. (2012a). Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. Neuron *75*, 133–142.

Livneh, U., and Paz, R. (2012b). Aversive-bias and stage-selectivity in neurons of the primate amygdala during acquisition, extinction, and overnight retention. J. Neurosci. *32*, 8598–8610.

Livneh, U., Resnik, J., Shohat, Y., and Paz, R. (2012). Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. Proc. Natl. Acad. Sci. USA *109*, 18956–18961.

Martin, R.F., and Bowden, D.M. (2000). Primate Brain Maps: Structure of the Macaque Brain. (Amsterdam: Elsevier Science).

Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature *459*, 837–841.

Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. Nat. Neurosci. *10*, 647–656.

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., Pitman, R.K., Ellis, C.B., Gold, A.L., Shin, L.M., Lasko, N.B., Zeidan, M.A., Handwerger, 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.

Morris, J.S., and Dolan, R.J. (2004). Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. Neuroimage *22*, 372–380.

Murray, E.A. (2007). The amygdala, reward and emotion. Trends Cogn. Sci. *11*, 489–497.

Ochsner, K.N., and Gross, J.J. (2005). The cognitive control of emotion. Trends Cogn. Sci. *9*, 242–249.

Pare, D., and Duvarci, S. (2012). Amygdala microcircuits mediating fear expression and extinction. Curr. Opin. Neurobiol. *22*, 717–723.

Paton, J.J., Belova, M.A., Morrison, S.E., and Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature *439*, 865–870.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat. Rev. Neurosci. *2*, 417–424.

Paz, R., and Pare, D. (2013). Physiological basis for emotional modulation of memory circuits by the amygdala. Curr. Opin. Neurobiol. *23*, 381–386.

Pearce, J.M., and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol. Rev. *87*, 532–552.

Pearce, J.M., and Mackintosh, N.J. (2010). Two theories of attention: a review and a possible integration. In Attention and Associative Learning: From Brain to Behavior, C.J. Mitchell and M.E. Le Pelley, eds. (Oxford: Oxford University Press).

Pitman, R.K., Rasmusson, A.M., Koenen, K.C., Shin, L.M., Orr, S.P., Gilbertson, M.W., Milad, M.R., and Liberzon, I. (2012). Biological studies of post-traumatic stress disorder. Nat. Rev. Neurosci. *13*, 769–787.

Preuschoff, K., and Bossaerts, P. (2007). Adding prediction risk to the theory of reward learning. Ann. N Y Acad. Sci. *1104*, 135–146.

Rauch, S.L., Shin, L.M., and Phelps, E.A. (2006). Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research—past, present, and future. Biol. Psychiatry *60*, 376–382.

Rescorla, R.A. (2007). Spontaneous recovery after reversal and partial reinforcement. Learn. Behav. *35*, 191–200.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black and W.F. Prokasy, eds. (New York: Appleton-Century-Crofts), pp. 64–99.

Resnik, J., Sobel, N., and Paz, R. (2011). Auditory aversive learning increases discrimination thresholds. Nat. Neurosci. *14*, 791–796.

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.

Roesch, M.R., Esber, G.R., Li, J., Daw, N.D., and Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. Eur. J. Neurosci. *35*, 1190–1200.

Rushworth, M.F., and Behrens, T.E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nat. Neurosci. *11*, 389–397.

Saleem, K., and Logothetis, N. (2007). A Combined MRI and Histology Atlas of the Rhesus Monkey Brain. (London: Elsevier).

Salzman, C.D., and Fusi, S. (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. Annu. Rev. Neurosci. *33*, 173–202.

Salzman, C.D., Paton, J.J., Belova, M.A., and Morrison, S.E. (2007). Flexible neural representations of value in the primate brain. Ann. N Y Acad. Sci. *1121*, 336–354.

Schiller, D., and Delgado, M.R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. Trends Cogn. Sci. *14*, 268–276.

Schiller, D., Levy, I., Niv, Y., LeDoux, J.E., and Phelps, E.A. (2008). From fear to safety and back: reversal of fear in the human brain. J. Neurosci. *28*, 11517–11525.

Schultz, W. (2012). Updating dopamine reward signals. Curr. Opin. Neurobiol. *23*, 229–238, http://dx.doi.org/10.1016/j.conb.2012.11.012, Published December 22, 2012.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Schultz, D.H., Balderston, N.L., and Helmstetter, F.J. (2012). Resting-state connectivity of the amygdala is altered following Pavlovian fear conditioning. Front. in Hum. Neurosci. *6*, 242.

Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J. Neurosci. *27*, 8366–8377.

Shin, L.M., and Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. Neuropsychopharmacology *35*, 169–191.

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.

Sotres-Bayon, F., and Quirk, G.J. (2010). Prefrontal control of fear: more than just extinction. Curr. Opin. Neurobiol. *20*, 231–235.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction. (Cambridge: MIT Press).

Tye, K.M., Cone, J.J., Schairer, W.W., and Janak, P.H. (2010). Amygdala neural encoding of the absence of reward during extinction. J. Neurosci. *30*, 116–125.

Wallis, J.D., and Kennerley, S.W. (2011). Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex. Ann. N Y Acad. Sci. *1239*, 33–42.

Wise, S.P. (2008). Forward frontal fields: phylogeny and fundamental function. Trends Neurosci. *31*, 599–608.