LETTERS

The primate amygdala represents the positive and negative value of visual stimuli during learning

Joseph J. Paton[1]*, Marina A. Belova[1]*, Sara E. Morrison[1] & C. Daniel Salzman[1,2,3]

Visual stimuli can acquire positive or negative value through their association with rewards and punishments, a process called reinforcement learning. Although we now know a great deal about how the brain analyses visual information, we know little about how visual representations become linked with values. To study this process, we turned to the amygdala, a brain structure implicated in reinforcement learning[1–5]. We recorded the activity of individual amygdala neurons in monkeys while abstract images acquired either positive or negative value through conditioning. After monkeys had learned the initial associations, we reversed image value assignments. We examined neural responses in relation to these reversals in order to estimate the relative contribution to neural activity of the sensory properties of images and their conditioned values. Here we show that changes in the values of images modulate neural activity, and that this modulation occurs rapidly enough to account for, and correlates with, monkeys' learning. Furthermore, distinct populations of neurons encode the positive and negative values of visual stimuli. Behavioural and physiological responses to visual stimuli may therefore be based in part on the plastic representation of value provided by the amygdala.

The complex anatomical connections of the amygdala, a collection of nuclei located deep in the medial temporal lobe, make it a prime candidate for providing a representation of the value of visual stimuli. The amygdala receives inputs from the visual system and from other sensory systems that represent reinforcing stimuli[6–8]. In addition, the amygdala is likely to receive error signals that represent stimuli in relation to expectations and that may be essential in creating an updated representation of value. The source of error signals for aversive learning has not been identified; however, midbrain dopamine neurons might supply such error signals for appetitive learning[9]. These signals could influence amygdala neural responses either directly[7] or indirectly through other brain structures such as the prefrontal cortex[6,10]. The convergence of information from each of these input pathways, perhaps combined with processing that occurs through intrinsic connections within the amygdala, may form a representation of visual stimulus value.

Unlike the anatomy of the amygdala, the physiological properties of amygdala neurons—especially with respect to learning the value of visual stimuli—remain poorly understood. We therefore recorded the activity of single amygdala neurons while monkeys learned the positive or negative value of new, abstract images during a trace-conditioning procedure[11] (Fig. 1a). In this task, a brief time interval (the trace interval) separated the presentation of a conditioned stimulus (CS) from an unconditioned stimulus (US). In each experiment, we assigned every CS a value: positive, negative or non-reinforced. After the image viewing and trace intervals, we delivered a liquid reward (after positive images), an aversive air-puff directed at the face (after negative images) or nothing (after non-reinforced images). The monkeys learned to associate each image with its value: during the trace interval they licked at a spout in anticipation of a liquid reward, or closed their eyes—a defensive behaviour—in anticipation of an air-puff. (We refer to the monkeys' eye closures as blinks, but the eyes typically stayed closed for longer than a blink, suggesting that this behaviour was purposeful.)

After the monkeys had learned the associations, without warning we reversed the values of the positive and negative images, and the monkeys learned the new values. Unlike a task in which subjects learn to avoid aversive USs, in our task, CSs predicted USs with 100% certainty. We tested every neuron during both appetitive and aversive conditioning for the same CS, allowing us to determine whether distinct amygdala neurons respond preferentially to positive or negative value for each CS. We report that amygdala neurons encode—in a flexible manner and in separate populations of neurons—the positive or negative value of conditioned images. Furthermore, these signals predict when monkeys learn, and develop rapidly enough to account for learning.

Anticipatory licking and blinking behaviour provided a measure of monkeys' learning about the positive and negative association of images. In the experiment shown in Fig. 2a, b, licking response rates were greater when an image was positive than when the same image was negative. Blinking responses showed the opposite trend. We used a change-point test to identify when the response rate changed significantly in relation to a value reversal ($P < 0.05$), represented graphically by a change in slope of the cumulative record of responses[12]. For image 1, the licking rate decreased significantly at trial 52 and blinking rate increased at trial 48 (22 and 18 trials after reversal, respectively). Because the monkey still licked sporadically when image 2 was negative, the change-point algorithm identified trial 29 as when licking began to increase, slightly before the reversal. Blinking decreased significantly starting at trial 35.

The primary objectives of our study were (1) to determine whether positive and negative conditioned visual stimuli engage the same or different populations of amygdala neurons, (2) to compare quantitatively the time course of how neural activity and behaviour change upon image value changes, and (3) to characterize the relative contributions of conditioned value and image identity to neuronal responses. In each of two monkeys (monkey V and monkey P), we used magnetic resonance imaging (MRI) to localize our recording electrode to the amygdala (Fig. 1b–d). We obtained complete data sets during monkeys' performance of initial and reversal learning for 196 neurons. To determine the extent to which amygdala neural activity reflected the conditioned values of images, or the images' sensory properties, we examined neural activity in relation to image value reversals. For the amygdala neuron recorded during the behaviour depicted in Fig. 2a, b, neural activity during the trace

[1]Center for Neurobiology and Behavior, [2]W. M. Keck Center on Brain Plasticity and Cognition, and [3]Department of Psychiatry, Columbia University, 1051 Riverside Drive, Unit 87, New York, New York 10032, USA.
*These authors contributed equally to this work.

interval was higher when images had a positive value than when the same images had a negative value (Fig. 2c–f), typical of a value-coding neuron. Other cells showed value-coding during the visual stimulus interval (see Supplementary Note 1 and Supplementary Fig. 1).

To characterize how quickly neural activity changed in relation to reversal of image value, we applied the change-point test[12] to trace interval neural responses (Fig. 2g, h). For image 1, neural activity began to decrease at trial 38 (8 trials after reversal), before either licking or blinking behaviour changed. For image 2, neural activity increased at trial 33, preceding the change in blinking behaviour. Overall, 100/196 (51%) of neurons changed activity in relation to value during the visual stimulus and/or trace intervals. For the cell shown in Fig. 2, a two-way analysis of variance (ANOVA) verified the influence of image value on neural responses (65% of variance accounted for by image value, $P < 0.05$), as well as a much smaller effect of image identity (8% of variance accounted for by image identity, $P < 0.05$). Across our population of neurons, image value often accounted for a statistically significant percentage of the variance (Fig. 3a, b; $P < 0.05$, two-way ANOVA). In addition, image identity had a significant effect in many cells, often overlapping with the representation of value at the single-cell level. The inset histograms show the difference in the strength of identity and value coding for all neurons with a significant effect of image identity, value or both factors. The two distributions are significantly different, with the trace interval showing significantly more value coding compared to image identity coding than the visual stimulus interval ($P < 10^{-5}$, Wilcoxon test; visual stimulus median difference −1.2; trace median difference 11.8). Thus, the representation of image identity weakens substantially after an image disappears, and neural activity then predominantly reflects the learned value of images.

To further characterize the representation of value in the amygdala, we determined quantitatively whether neurons responded more strongly when an image had a positive or negative value. We used a receiver operating characteristic (ROC) analysis[13] to estimate the extent to which activity before and after an identified change point differed (Fig. 3c). Each data point from this analysis reflects the change in activity for a single image in a single interval. By convention, we used the negative value responses as the reference distribution, so ROC values >0.5 indicated cells that responded more when an image was positive than when the same image was negative. All ROC values except two were significantly different from 0.5 ($P < 0.05$, permutation test). The bi-modally distributed data indicate that some amygdala neurons prefer negative conditioned stimuli, and others prefer positive conditioned stimuli (see Supplementary Note 2).

The activity of neurons encoding positive and negative values was not related to motor responses (licking and blinking; see Supplementary Figs 3, 4). Moreover, the development of value coding was not an artefact of using different modalities for our reinforcing stimuli (although the air-puff and liquid reward both have auditory and somatosensory components). Many neurons encoding positive image value responded to air-puffs, and many neurons whose activity reflected negative image value responded to rewards (Table 1 and Supplementary Fig. 5). Thus, neurons encoding value received information about both kinds of reinforcing stimuli. Finally, neurons representing positive and negative values did not demonstrate clear anatomical clustering (Fig. 1e, f).

We next characterized how neurons represent value across time during trials. An ROC analysis applied to consecutive overlapping time-windows of 100 ms (advanced in steps of 20 ms), revealed that different neurons encoded image value during different overlapping time segments (Fig. 3d). Some neurons represented value exclusively during the visual stimulus or trace intervals, but other cells sustained

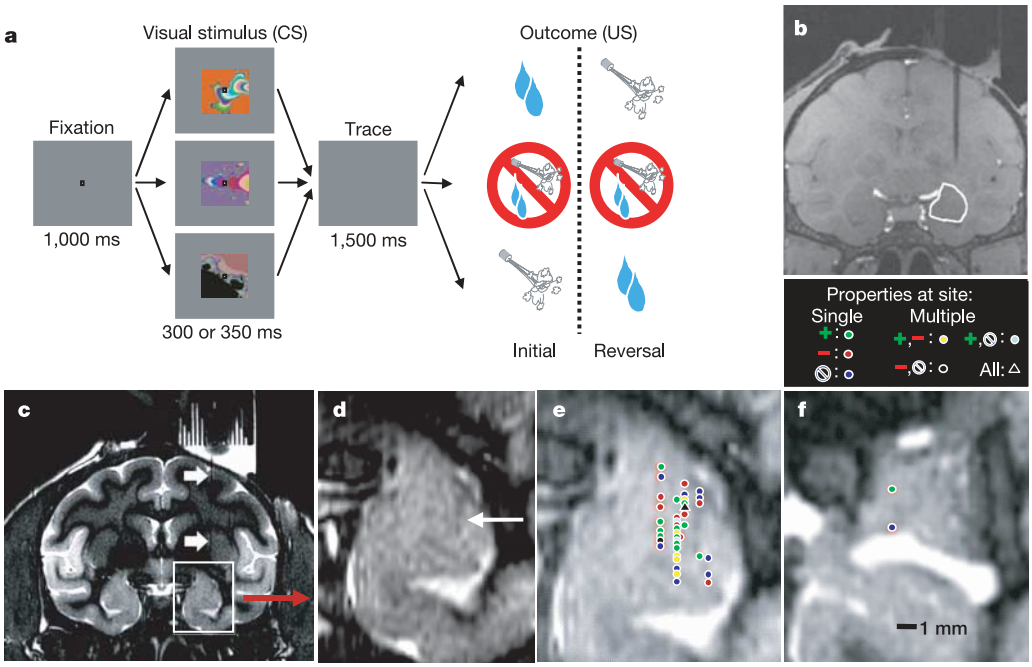


Figure 1 | Task and brain MRI. **a**, Sequence of events for the three trial types. Top and bottom squares, image values reverse. Middle square, image always non-reinforced. **b**, Coronal MRI acquired with a two-dimensional (2D) spoiled gradient recalled acquisition (SPGR) sequence in monkey V, showing the artefact from a tungsten microelectrode dorsal to the amygdala (outlined in white). **c**–**f**, Coronal MRI with 2D inversion recovery (IR) sequence (**c**, arrows point to the electrode). Magnified images show the probable border of the lateral nucleus (arrow in **d**) and recording site locations (slice in **f** is immediately posterior to **e**). We collapsed recording sites spanning 2 mm in the anterior–posterior dimension onto each image slice, in many cases resulting in the superposition of multiple cells with different properties (see key above **f** for symbols denoting properties; '+' denotes positive value-coding, '−' denotes negative value-coding, and 'no' symbol indicates no value-coding). Recording sites from monkey P occurred in an overlapping region of the amygdala.

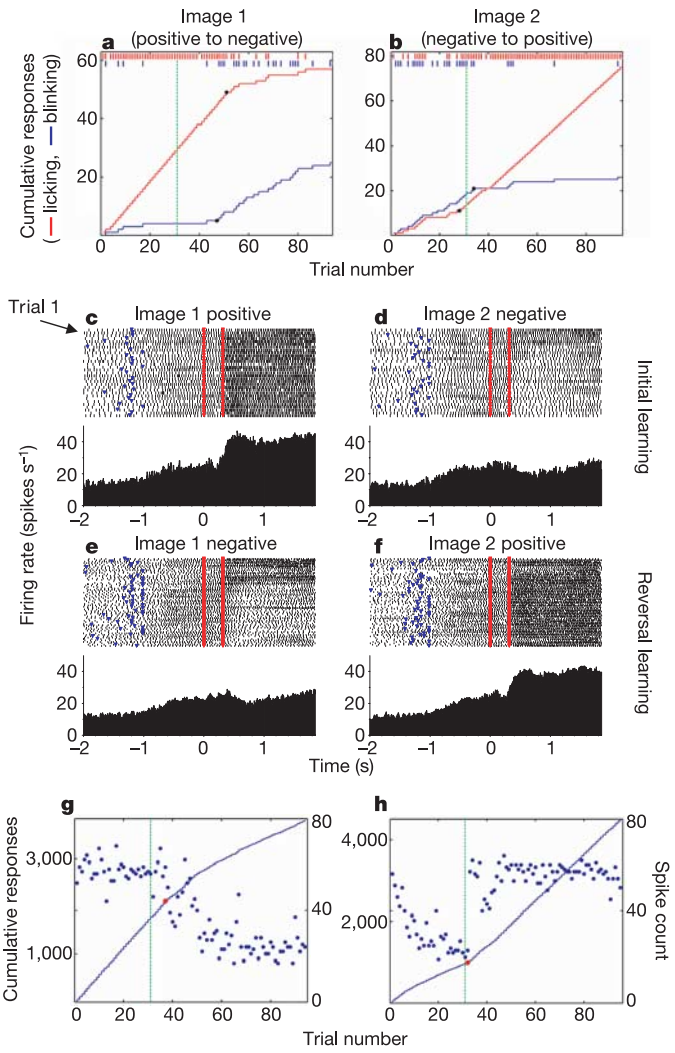


Figure 2 | Behaviour and neural activity from a single amygdala neuron during learning. a, **b**, Cumulative (curves) and trial-by-trial (tick marks) measures of licking (red) and blinking (blue), plotted as a function of trial number for images 1 and 2. Black dots indicate change points. Vertical green lines indicate value reversals. **c**–**f**, Rasters and peri-stimulus time histograms (PSTHs), truncated at US delivery, for the amygdala cell recorded during the same experiment. Each row of dots represents the timing of action potentials during one trial. PSTHs sum activity across trials and were smoothed by taking a 10-ms moving average of activity. Blue ticks indicate fixation point onset. Red ticks indicate visual stimulus onset/offset. **g**, **h**, Spike count and cumulative spike count during the trace interval, plotted as a function of trial number for images 1 and 2. Red dots indicate change points.

a representation across some or all of both intervals. Across the population of neurons, signals representing value developed rapidly after visual stimulus onset and were maintained through the trace interval (Fig. 3d; for positive-coding cells, the first bin significantly greater than 0.5 occurred 80–180 ms after visual stimulus onset, $P < 0.05$ by t-test; for negative-coding cells, the first significant bin was 120–220 ms after visual stimulus onset). This temporally extended representation of value may be useful for the development of adaptive behaviour—such as licking and blinking—that anticipates reinforcement. Furthermore, the representation of value over time could have an important role in reinforcement learning: computations of error signals may depend on a comparison of neural signals that represent value as a function of time[9].

If decisions to lick or blink in anticipation of a reward or punishment are based in part on the representation of value we describe in the amygdala, then, on average, neural activity should change at the same time as learning occurs. To compare behavioural learning of the image value reversals to changes in neural activity, we plotted the change-point for neural activity against the trial when behaviour changed (Fig. 4a, b; each data point compares a neural activity change point with either a blinking or licking change point; the distributions of licking and blinking change points were not significantly different from one another; $P > 0.25$, t-test). Learning was significantly correlated with changes in neural activity during both the visual stimulus and trace intervals (visual stimulus interval: monkey V, $P = 0.02$, $r^2 = 0.06$; monkey P, $P < 10^{-5}$, $r^2 = 0.44$; trace interval: monkey V, $P < 10^{-5}$, $r^2 = 0.40$; monkey P, $P < 10^{-5}$, $r^2 = 0.32$). The difference between neural and behavioural change points was near zero (Fig. 4 histograms; monkey V: visual stimulus mean difference 1.4 trials, trace mean difference −0.8 trials; monkey P: visual stimulus mean difference −0.5 trials, trace mean difference 0.3 trials). For all comparisons, the onsets of changes in behavioural and neural responses were statistically indistinguishable (paired t-test, $P > 0.05$). In addition, the distributions comparing change points during the visual stimulus and trace intervals were not significantly different ($P > 0.1$, t-test).

These results indicate that amygdala neurons can rapidly adjust their activity to reflect image value; these changes in activity both predict when monkeys learn and develop, on average, as fast as monkeys learn. The fact that some cells change their activity more rapidly than monkeys learn suggests that monkeys cannot selectively 'listen' to these neurons, or they would demonstrate faster learning. Learning may be better accounted for if monkeys must evaluate signals from the population of neurons, which includes neurons that change their activity more slowly.

The data presented so far compare the onset of changes in behavioural and neural responses. We next compared the time course of average neural and behavioural responses across the 100 neurons encoding image value (as identified by the change-point test). We normalized and then averaged neural activity and behaviour from the 20 trials before and after the value reversal of each image. To describe the behavioural and neural changes around reversals, we fitted the behavioural and neural data with sigmoid functions (Fig. 4c, see Supplementary Fig. 6). The 95% prediction intervals for these functions overlapped, indicating that the trajectories of changes in behavioural responses and neural activity were similar under this model. In contrast, the same analysis on the non-reinforced images for the same cells showed no change in activity or behaviour in relation to the reversal of the positive and negative images (Fig. 4d). These data demonstrate that, on average, neurons begin to change their activity within very few trials of a change in image value, and

Table 1 | Responses to reinforcing stimuli in amygdala neurons that code value

	Excitatory response to air-puff	Inhibitory response to air-puff	Excitatory response to reward	Inhibitory response to reward	No response to reward or air-puff
Cells coding positive CS value ($n = 61$)	34	12	23	6	9
Cells coding negative CS value ($n = 39$)	18	8	11	7	7

Excitatory and inhibitory responses to air-puffs and rewards for all cells encoding the positive or negative value of images. The number of cells adds up to more than the number of value-coding neurons because many cells were modulated by both air-puff and reward.

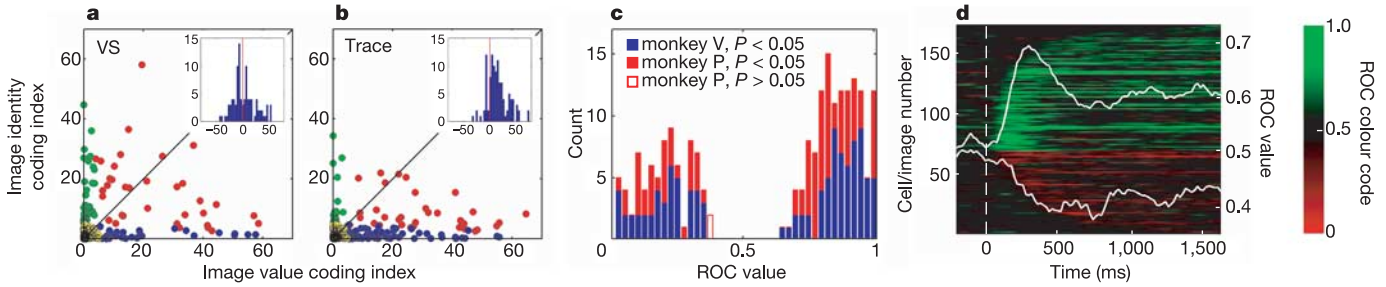


Figure 3 | Amygdala neurons encode the positive and negative value of visual stimuli. **a**, **b**, Image identity coding index (II_CI) plotted against image value coding index (IV_CI) for the visual stimulus (VS) and trace intervals. Green, blue and red dots, $P < 0.05$ for II_CI, IV_CI or both indices, respectively (two-way ANOVA). Yellow dots, not significant. Inset histograms show II_CI subtracted from IV_CI for non-yellow data points. See Supplementary Fig. 9. **c**, Encoding of positive and negative value shown by ROC analysis. **d**, Development of value signals as a function of time. Each row in the colour map represents value coding for a neuron during presentation of a single image, with positive and negative cell rows sorted in opposite order according to the first post-visual-stimulus data point significantly different from 0.5 ($P < 0.05$, permutation test). White curves show mean ROC values across the positive- and negative-coding populations. Time 0 is the start of the bin spanning from 0–99 ms after visual stimulus onset.

that—across the population of neurons—the rate of changes in activity closely tracks behavioural learning.

The rapid appearance of a representation of visual stimulus value in the amygdala raises the possibility that this representation is formed locally. If a representation of value is first formed elsewhere in the brain, it would have to provide signals that represent value at earlier time points within trials, and activity would have to change even more rapidly across trials compared to what we observe in the amygdala. Moreover, the correlation between changes in amygdala activity and behaviour strongly suggests that the amygdala representation is used by the monkey, regardless of its source. Nonetheless, other brain structures such as the orbitofrontal cortex probably have critical roles in learning on our task. The amygdala seems to update its representation as quickly as the orbitofrontal cortex changes its activity during reversal learning (using a different task, in which monkeys learn to avoid aversive stimuli, and different analytic methods)[14]. Other studies that have not used aversive stimuli also suggest that the orbitofrontal cortex provides a representation of reward value[15–17]. Understanding the complex interactions between the amygdala and orbitofrontal cortex (and other structures) will require substantial investigative effort in order to tease apart the distinct contributions of each brain area to reinforcement learning. In rodents, some evidence suggests that the physiological properties of amygdala and orbitofrontal neurons are interdependent[18,19].

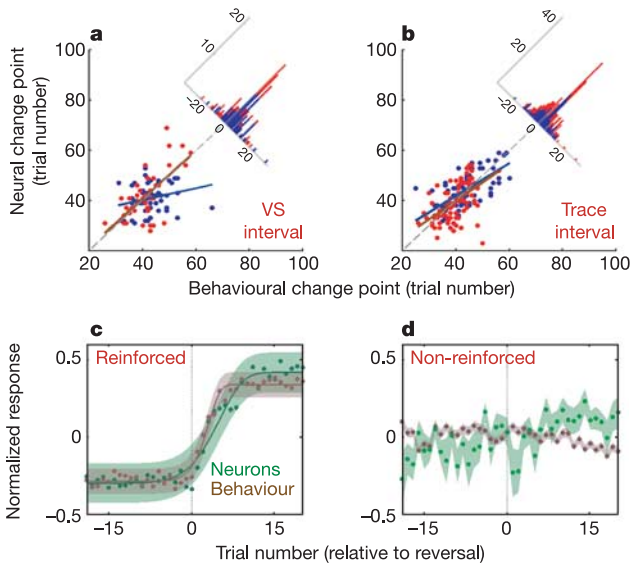


Figure 4 | The relationship between changes in neural activity and behavioural responses. **a**, **b**, Onset of changes in neural activity plotted against the onset of changes in licking or blinking for visual stimulus (VS) and trace interval activity. Histograms show the difference between neural and behavioural change points. Blue, data and regression lines for monkey V; red, data and regression lines for monkey P. **c**, Average normalized neural activity and behavioural responses plotted as a function of trial number relative to the reversal in image value. Shaded regions indicate 95% prediction intervals for best-fit Weibull functions. **d**, Similar representation as in **c**, applied to the non-reinforced images. Shaded regions show s.e.m. of data points.

Human and non-human primate behaviour is strongly influenced by visual processing and, therefore, the assignment of value to visual stimuli is a fundamental process that underlies many of our actions. In lower mammals, previous studies have investigated amygdala neurophysiology in relation to learning; however, these studies have used auditory or olfactory conditioned stimuli, and tasks that either do not include appetitive conditioning or that involve learning to avoid an aversive unconditioned stimulus[1,20]. The primate amygdala has extensive interconnections with the visual cortex[21], and previous lesion studies in monkeys and neuroimaging studies in humans have suggested that the amygdala is involved in updating a representation of the value of visual stimuli[3,22,23] (see Supplementary Note 3). Previous neurophysiological recordings in monkey amygdala have used different methodological procedures than those here, and provide conflicting data regarding whether the amygdala contains an updated representation of visual stimulus value[24–26].

Our data establish that distinct amygdala neurons encode the positive and negative learned value of images. The representation of value is often linked to image selectivity initially, but this representation becomes transformed over time so that value, and not image selectivity, is predominantly encoded after visual stimuli disappear. Signals that reflect value develop rapidly, and evolve as quickly, on average, as monkeys learn. Furthermore, the timing of changes in amygdala activity predicts when monkeys learn. Thus, in principle, decisions to either lick or blink during performance of our task could be based on the plastic representation of visual stimulus value provided by the amygdala.

Reinforcement learning, a process by which sensory stimuli become associated with positive or negative values, is intimately related to physiological and behavioural responses thought to represent emotions, because such responses frequently occur as a reaction to sensory stimuli endowed with value. Amygdala neural signals that reflect CS value may influence a variety of interconnected brain structures—including the ventral striatum, prefrontal cortex, sensory cortices, the hippocampus and rhinal cortices, and subcortical brain areas—involved in aspects of emotional responses and reinforcement learning. Different amygdala nuclei have complex

patterns of connectivity with each of these structures[6,7,10,21,27], suggesting that neural signals in the amygdala may be used for very different purposes depending upon their anatomical destination. For example, signals that reflect CS value may influence attention[28] or computations of error signals[9] through projections from the amygdala central nucleus to the ventral tegmental area and substantia nigra[29]. In addition, value signals could influence behavioural and physiological responses to visual stimuli through other subcortical projections from the central nucleus[30], and through projections from the basal and accessory basal nuclei to the prefrontal cortex and ventral striatum[6,7,10]. The amygdala neural signals we describe—which reflect the learned value of visual stimuli—are therefore only one component of the intricate neural circuitry that mediates reinforcement learning and our emotional lives. To understand how this circuitry operates, future studies must identify how dynamic interactions between the amygdala and interconnected brain structures function to support a wide range of emotional and cognitive behaviours.

METHODS

Behavioural task. In every experiment, monkeys learned the positive, negative or non-reinforced value of three new images through a trace-conditioning procedure. In individual trials, monkeys centred their gaze at a fixation point (FP) for 1 s and viewed an image for 300 or 350 ms, determined empirically for each monkey to minimize broken fixation behaviour (see Supplementary Fig. 7). After a 1,500-ms trace interval during which monkeys were free to move their eyes, we delivered a US (for positive images, a 0.2–0.9 ml liquid reward; for negative images, a 50–100 ms, 40–60 psi aversive air-puff directed at the face; for non-reinforced images, nothing). We reversed image value assignments without warning at least 8 trials after monkeys learned the initial contingencies, as assessed by viewing licking and blinking data during the experiment. All trials were pseudorandomly interleaved. Images were fractal patterns constructed with a custom-written Matlab (Mathworks) program. Images had no inherent value and were easily distinguishable from each other and from recently shown images on previous days.

Data collection. We recorded neural activity from the right amygdala of two rhesus monkeys (*Macaca mulatta*) weighing 4–10 kg. All animal procedures conformed to NIH guidelines and were approved by the Institutional Animal Care and Use Committees at New York State Psychiatric Institute and Columbia University. We positioned recording chambers based on anatomical information acquired using MRI. Conventional metal microelectrodes (1–4) (FHC Instruments) were individually controlled and advanced into the brain through dura-puncturing guide tubes using either a motor-controlled hydraulic microdrive (Narishige) or a motorized multi-electrode drive (NAN). Signal amplification, filtering, digitizing of spike waveforms, and spike-sorting using a principal component analysis platform (online, with offline verification) were accomplished using the Plexon system. While we searched for neurons, monkeys performed either no task or a fixation task without images present. We studied all well-isolated neurons (196 neurons in total: 92 from monkey V, 104 from monkey P). We reconstructed recording sites using MRI combined with laboratory notes. Neurons were located in the middle 50% of the anterior–posterior extent of the amygdala, in the lateral 75% of the medial–lateral extent, and in the dorsal 75% of the dorsal–ventral extent. All neurons were probably located in either the basal, accessory basal, lateral or central nucleus, or the intercalated masses.

Licking was measured by the interruption of an infrared beam by the tongue extending to the reward tube, and blinking by the loss in signal from an infrared eye tracker (ASL), which corresponded to blinks as verified by a video camera feed. Licks or blinks occurring during the 500 ms preceding US delivery were scored as responses.

Data analysis. Based on an analysis of visual response latency (see Supplementary Fig. 8), we divided the trial into visual stimulus (90 ms after image appearance until 90 ms after image disappearance) and trace (90 ms after image disappearance until US delivery) intervals. Our data analysis had four principal goals. (1) To find the trial closest to image value reversals at which neural and behavioural (licking, blinking) responses began to change, we used a change-point test[12] ($P < 0.05$, correcting for multiple comparisons). (2) To characterize the relative contribution of image identity (that is, the sensory properties of images) and value on neural activity, we performed a two-way ANOVA. We defined three indices, corresponding to the percentage of variance accounted for by image identity, image value and their interaction (image identity, image value, and image value–identity interaction coding indices). (3) To determine whether amygdala neurons responded more when an image had a positive or negative value, we compared spike counts from the 20 trials on either side of an identified change point using a receiver operating characteristic (ROC) analysis[13] on neurons categorized as value-coding (209 intervals from 100 cells coding image value during the visual stimulus, trace or both intervals; 72/100 cells changed activity in opposite directions for both images upon reversal; 28/100 neurons changed responses for one image, but two-way ANOVA showed significant effects of image value). We also characterized the development within trials of signals representing value, computing an ROC value in 100-ms windows moved in 20-ms steps across the trial (performed on 172 image/cell combinations from the 100 value-coding cells). (4) To compare the trajectories of changes in neural and behavioural responses, we normalized and averaged neural activity and behaviour from each of the 20 trials before and after image value reversal across the 100 value-coding neurons. We fitted the behavioural and neural data with a Weibull function:

$$f(x) = u + (1 - l - u)\exp(-x/\alpha)^{\beta} \quad (1)$$

which modelled normalized response as a function of trial number. We performed a similar normalization and averaging procedure for non-reinforced trials, analysing the 20 trials before and after the trial that the other images reversed. Details of all methods are described further in Supplementary Methods.

Received 20 September; accepted 25 November 2005.

1. Maren, S. & Quirk, G. J. Neuronal signaling of fear memory. *Nature Rev. Neurosci.* **5,** 844–852 (2004).
2. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23,** 155–184 (2000).
3. Baxter, M. & Murray, E. A. The amygdala and reward. *Nature Rev. Neurosci.* **3,** 563–573 (2002).
4. Everitt, B. J., Cardinal, R. N., Parkinson, J. A. & Robbins, T. W. Appetitive behaviour: impact of amygdala-dependent mechanisms of emotional learning. *Ann. NY Acad. Sci.* **985,** 233–250 (2003).
5. Holland, P. C. & Gallagher, M. Amygdala circuitry in attentional and representational processes. *Trends Cogn. Sci.* **3,** 65–73 (1999).
6. Stefanacci, L. & Amaral, D. G. Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. *J. Comp. Neurol.* **451,** 301–323 (2002).
7. Amaral, D., Price, J., Pitkanen, A. & Carmichael, S. in *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (ed. Aggleton, J.) 1–66 (Wiley-Liss, New York, 1992).
8. McDonald, A. J. Cortical pathways to the mammalian amygdala. *Prog. Neurobiol.* **55,** 257–332 (1998).
9. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
10. Ghashghaei, H. T. & Barbas, H. Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* **115,** 1261–1279 (2002).
11. Pavlov, I. P. *Conditioned Reflexes* (Oxford Univ. Press, London, 1927).
12. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proc. Natl Acad. Sci. USA* **101,** 13124–13131 (2004).
13. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (John Wiley and Sons, New York, 1966).
14. Rolls, E. T., Critchley, H. D., Mason, R. & Wakeman, E. A. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J. Neurophysiol.* **75,** 1970–1981 (1996).
15. Wallis, J. D. & Miller, E. K. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* **18,** 2069–2081 (2003).
16. Roesch, M. R. & Olson, C. R. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304,** 307–310 (2004).
17. Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* **398,** 704–708 (1999).
18. Saddoris, M. P., Gallagher, M. & Schoenbaum, G. Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. *Neuron* **46,** 321–331 (2005).
19. Schoenbaum, G., Setlow, B., Saddoris, M. P. & Gallagher, M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39,** 855–867 (2003).
20. Schoenbaum, G., Chiba, A. & Gallagher, M. Orbitalfrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neurosci.* **1,** 155–159 (1998).
21. Freese, J. L. & Amaral, D. G. The organization of projections from the amygdala to visual cortical areas TE and V1 in the macque monkey. *J. Comp. Neurol.* **486,** 295–317 (2005).
22. Gottfried, J., O'Doherty, J. & Dolan, R. J. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* **301,** 1104–1107 (2003).
23. LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E. & Phelps, E. A. Human

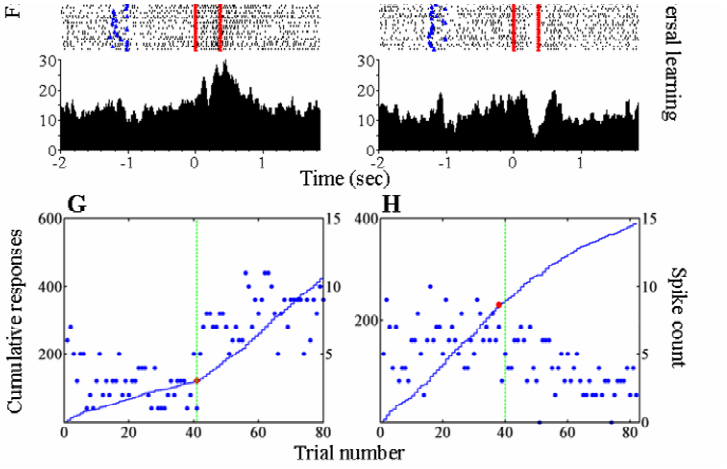
amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* **20,** 937–945 (1998).
24. Sanghera, M. K., Rolls, E. T. & Roper-Hall, A. Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Exp. Neurol.* **63,** 610–626 (1979).
25. Rolls, E. in *The Amygdala: A Functional Analysis* (ed. Aggleton, J.) 447–478 (Oxford Univ. Press, New York, 2000).
26. Nishijo, H., Ono, T. & Nishino, H. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J. Neurosci.* **8,** 3570–3583 (1988).
27. Stefanacci, L., Suzuki, W. A. & Amaral, D. G. Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. *J. Comp. Neurol.* **375,** 552–582 (1996).
28. Lee, H. J. *et al.* Role of amygdalo-nigral circuitry in conditioning of a visual stimulus paired with food. *J. Neurosci.* **25,** 3881–3888 (2005).
29. Fudge, J. L. & Haber, S. N. The central nucleus of the amygdala projection to dopamine subpopulations in primates. *Neuroscience* **97,** 479–494 (2000).
30. Davis, M. in *The Amygdala: A Functional Analysis* (ed. Aggleton, J.) 213–287 (Oxford Univ. Press, New York, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. R. Gallistel for discussions and for assistance with the change-point test; S. Dashnaw and J. Hirsch for MRI support; C. A. Mason, E. R. Kandel, M. N. Shadlen and members of the Mahoney Center at Columbia for comments on the manuscript; and M. E. Goldberg, J. E. LeDoux and W. T. Newsome for mentoring during a career development award to C.D.S. This work was supported by the Keck foundation, grants from the NIMH, and the Klingenstein, Sloan and NARSAD foundations, and by a Charles E. Culpeper Scholarship award from Goldman Philanthropic Partnerships to C.D.S. J.J.P. received support from NICHD and NEI institutional training grants. S.E.M. received support from an NSF graduate research fellowship.

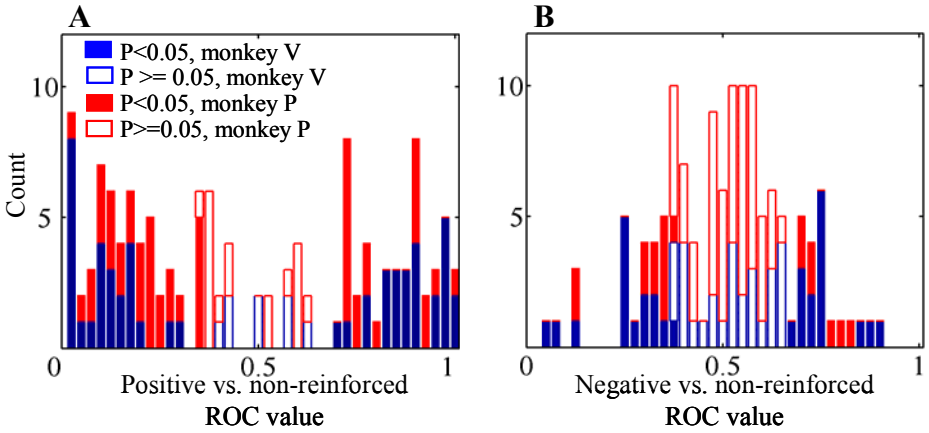
Author Contributions J.J.P. and M.A.B. performed all experiments and conducted data analyses. S.E.M. performed some of the data analyses and contributed to many discussions. Experiments were designed and implemented in the laboratory of C.D.S.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D.S. (cds2005@columbia.edu).



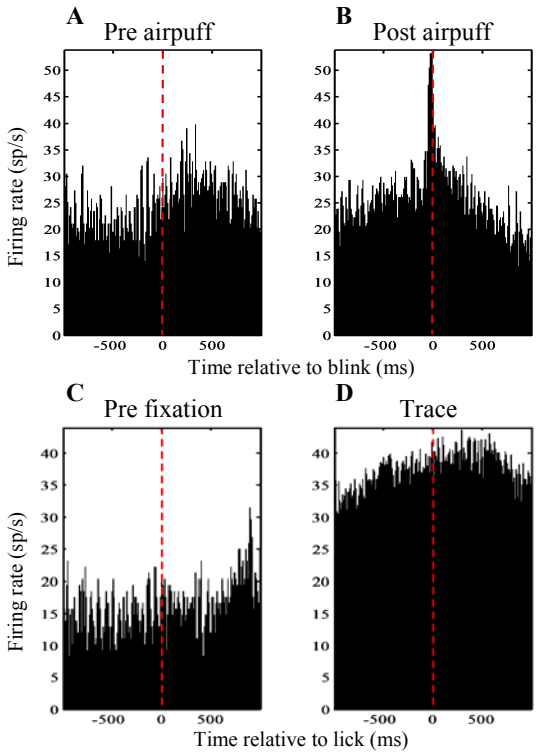
Supplementary Figure 1. Example of an amygdala neuron whose activity reflects value during the visual stimulus interval. This cell responded more strongly when an image was negative than when the same image was positive. **A,B.** Licking and blinking responses (red and blue tick marks, respectively, at the top of the panels), along with cumulative plots of responding, plotted as a function of trial number for both images that underwent a reversal. The black dots indicate the change points for licking and blinking. The change point marks the trial where the slope changes significantly on each cumulative plot. **C-F.** Rasters and PSTHs for the amygdala cell recorded during the learning depicted in **A**,**B**. Left to right tick marks show fixation point onset (blue), and visual stimulus onset/offset (red ticks, aligned in time); the raster is truncated at US delivery. **G,H.** Spike count and cumulative spike count (from the visual stimulus interval) plotted as a function of trial number for Image 1 (**G**) and Image 2 (**H**) plotted continuously across the reversal. Red dots: activity change point, where the slope of the cumulative plots changes. Green vertical line, reversal trial.

Supplementary Figure 2, Salzman

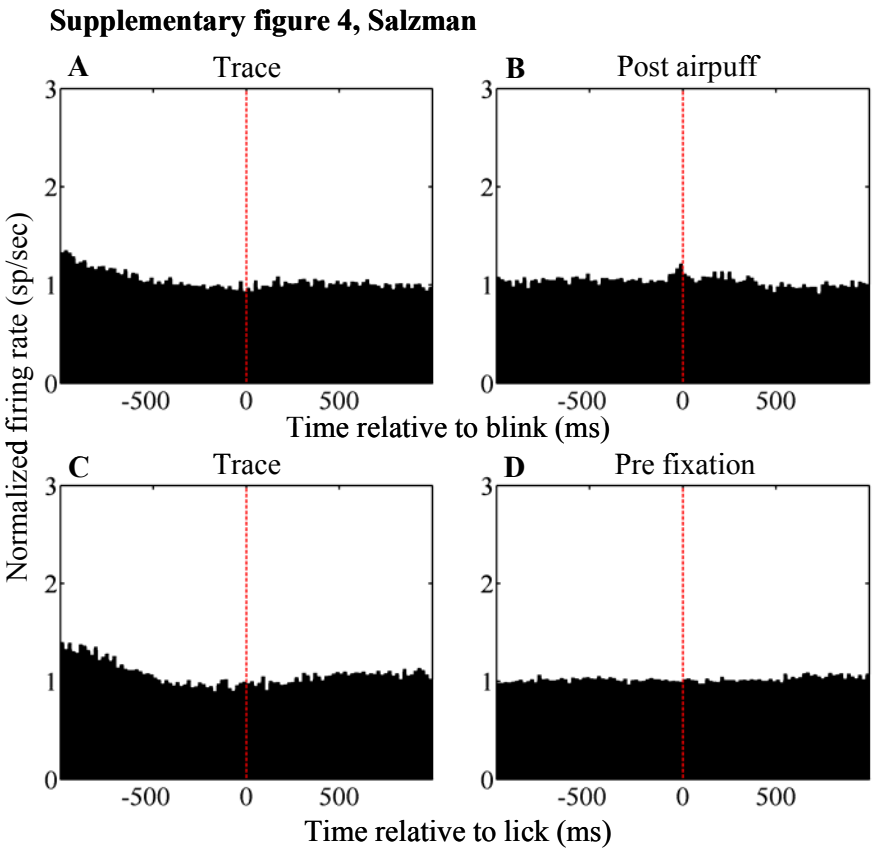


Supplementary Figure 2. Comparison of trace interval activity from non-reinforced images to activity from positive (**A**) and negative (**B**) images. We performed a similar analysis to that shown in Fig. 3C, but which instead compared responses on non-reinforced trials to responses on positive or negative trials in a time period from 350 ms after visual stimulus offset until the time of US delivery (this time period was chosen to eliminate any possibility of a visual response). We selected the cells that responded differentially to positively and negatively conditioned images during the trace interval, excluding cells in which behavior on non-reinforced trials was not different from positive or negative trials (typically, both licking and blinking rates are reduced on non-reinforced trials). We then compared activity from the 20 trials before and after the change in activity related to image value reversal to activity from the non-reinforced image trials interleaved during the same time period. We used the non-reinforced trials as the reference distribution for both comparisons. Each cell contributes up to 2 data points (1 data point per image). **A.** Positive image responses compared to non-reinforced image responses. 104/130 data points had ROC values significantly different from 0.5 (permutation test, $p < 0.05$). **B.** Negative image responses compared to non-reinforced image responses. 47/129 data points have ROC values significantly different from 0.5 (permutation test, $p < 0.05$). We did not perform a similar analysis on visual stimulus activity because those responses are confounded by image selectivity (see Figure 3A).

Supplementary figure 3, Salzman

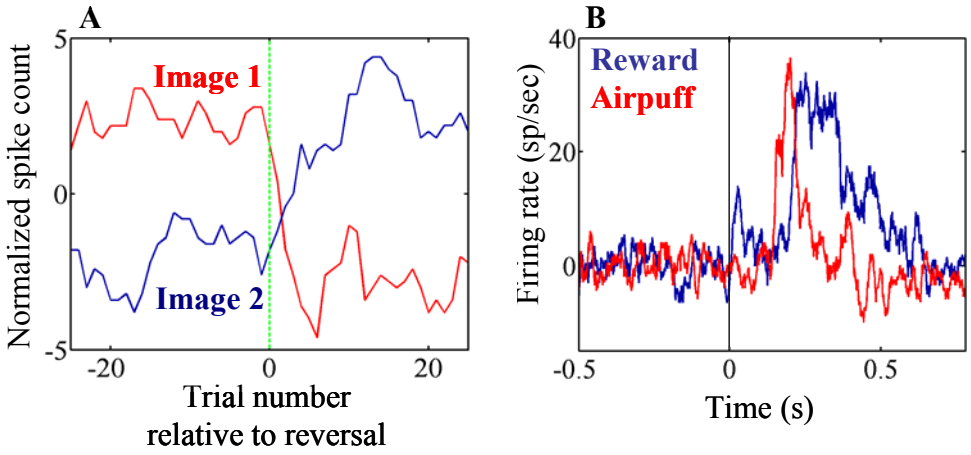


Supplementary Figure 3. Amygdala neural activity is not specifically related to motor responses (licking and blinking) for the cell depicted in Fig. 2. **A,B.** Blink-triggered spike histograms. Blinks were sorted into two categories depending on whether they occurred during the one second before airpuff delivery or the one second following airpuff delivery, and neural activity was aligned on these events. In **B**, the peak in activity just prior to blink onset was a sensory response to air puff delivery, not a response to blinking itself. The peak in activity does not precede blinks that occur during the trace interval (**A**). **C,D.** Lick-triggered spike histograms. Licks were sorted into two categories depending on whether they occurred in the one second before fixation point onset, or the last one second of the trace interval, and neural activity was aligned on these events. In all cases, spikes were binned into 10 ms bins and summed, before smoothing with a 2 bin moving average. The vertical dashed red line marks the time of blink or lick onset, respectively. The cell shows no stereotyped response around the two behaviors, indicating that changes in amygdala responses with learning cannot be attributed to motor signals used for licking and blinking. We performed similar analyses on all the cells in our sample. We used the same latency analysis described in the methods section to look for a response in the 200 ms period preceding a blink or an airpuff compared to a "background" level of activity in the period preceding blinks or licks by 200 – 400 ms. We conducted this analysis separately for blinks and licks occurring during the last 1 sec of the trace interval, and those responses occurring at other times during the trial. We did not find evidence for a consistent motor-related signal in any cell.

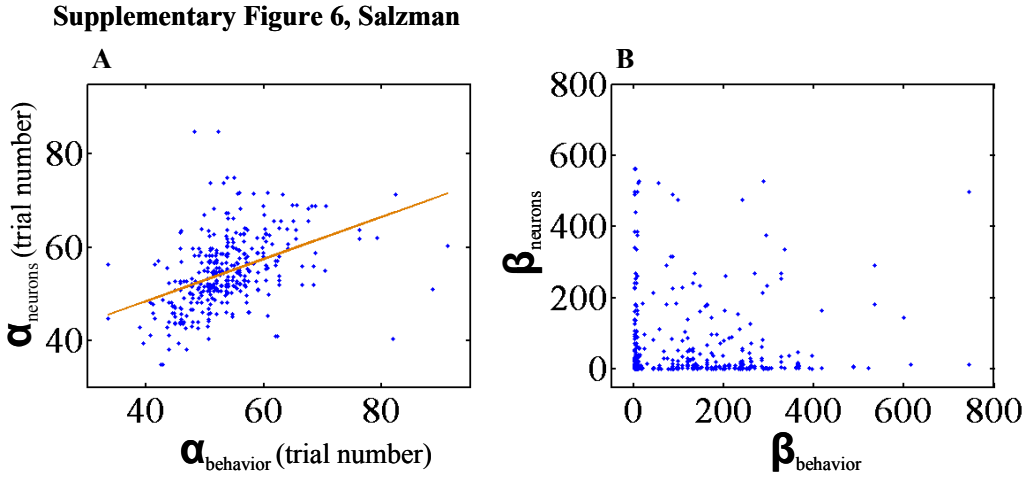


Supplementary Figure 4. Amygdala neural activity is not specifically related to motor responses (licking and blinking) across the population of cells recorded. **A,B.** Blink-triggered spike histograms. Blinks were sorted into two categories depending on whether they occurred during the one second before air puff delivery (trace interval) or the one second following air puff delivery. Neural activity was aligned on these events, normalized by dividing by the median response, and averaged across all neurons. There are no consistent peaks in activity occurring in relation to blinks during these time intervals across the population of neurons. The peak in activity seen before blinks occurring within 1 sec of an air puff is not present in advance of blinks occurring in the 1 sec before air puff delivery. The peak in activity seen for blinks after air puffs is due to a sensory response to the air puff itself that we commonly observe (see Table 1). **C,D.** Lick-triggered spike histograms. Licks were sorted into two categories depending on whether they occurred in the one second before fixation point onset, or the last one second of the trace interval. Neural activity was aligned on these events, normalized by dividing by the median response, and averaged across all neurons. Again, the analysis reveals that there are no consistent peaks in activity in relation to licks for the different time intervals across experiments. For all plots above, spikes were binned into 10 ms bins and summed, before smoothing with a 2 bin moving average. The vertical dashed red line marks the time of blink or lick onset, respectively. Overall, cells show no stereotyped response around the two behaviors, indicating that changes in amygdala responses with learning cannot be attributed to motor signals used for licking and blinking.

Supplementary Figure 5, Salzman

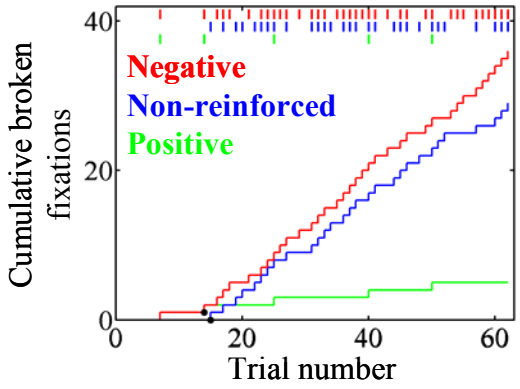


Supplementary Figure 5. Value coding is not simply due to cells receiving information from only reward pathways (for positive cells), or from only pathways representing air puffs (for negative cells). **A.** Normalized neural activity during the visual stimulus interval, smoothed by taking a 5 trial moving average, plotted as a function of trial number for 2 images in one experiment. One image started with a positive value and changed to negative, and the other image started with a negative value and reversed to positive. We normalized activity by subtracting the mean response of each image. This cell responded more when an image was positive than when the same image was negative. The green line demarcates when reversal occurs. **B.** PSTH aligned on the time of reinforcement (time 0) for negative trials (red line; air puff delivered) and positive trials (blue line; reward delivered), with baseline activity from the trace interval subtracted. Although this cell coded the positive value of images, it responded strongly to both air puff and reward.



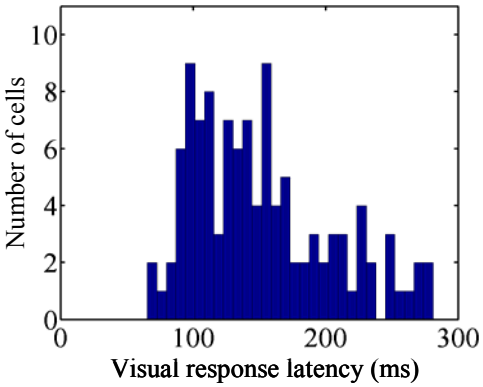
Supplementary Figure 6. The relationship between the trajectories of changes in neural activity and behavioral learning on a cell by cell basis. In addition to fitting sigmoid functions to the average normalized data across our population of value-coding cells, we also fit the same functions to data from the 20 trials before and after reversal for each experiment. Each value-coding neuron contributed up to 4 data points for this analysis (1 for each image from the visual stimulus and trace interval). **A.** α from the fit to behavioral data plotted against α from the fit to neural data. The estimate of α for neurons and behavior was significantly correlated across experiments ($p < 0.0001$, $r^2 = 0.16$), just as the onset of changes in neural activity and behavior were correlated. For this plot, we excluded 69 data points where the estimate of α was greater than 100, which corresponded to cases where either behavioral or neural changes occurred long after reversal, or when data was particularly noisy. **B.** β from the fit for behavior plotted as a function of β from the fit of neural activity. The data show that many β estimates have quite high values, corresponding to steep slopes in the fit sigmoids. We frequently observed this phenomenon when changes in behavior or neural activity appeared to be like a step function.

Supplementary Figure 7, Salzman



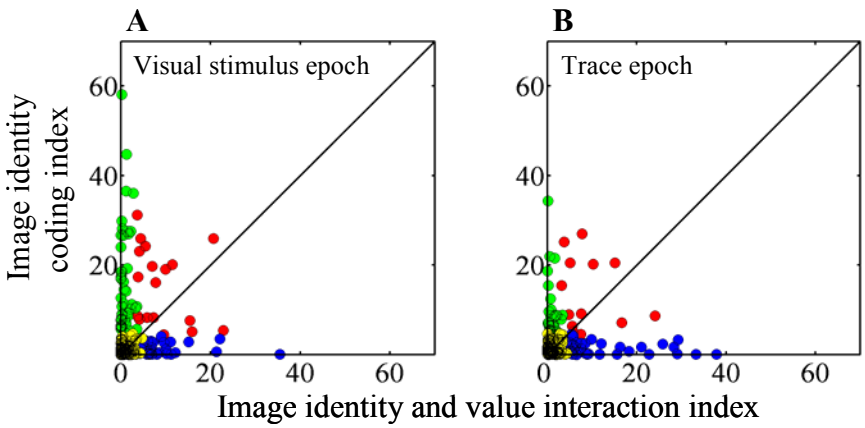
Supplementary Figure 7. Broken fixation behavior increases when a monkey learns the negative or non-reinforced value of a visual stimulus. For this experimental session, we presented the visual stimulus for 1 sec, and broken fixation behavior typically occurred 400 – 750 ms after image onset. Broken fixation and cumulative broken fixations plotted as a function of trial number for positive (green), non-reinforced (blue), and negative (red) image trials. The tick marks indicate trials where the monkey broke fixation. A significant increase in broken fixations occurs after 15 trials for non-reinforced images, and after 14 trials for negative trials ($p < 0.05$, change point test). These data demonstrate that the monkey tries to avoid the non-reinforced and negative trials by breaking fixation.

Supplementary Figure 8, Salzman



Supplementary Figure 8. Histogram of the visual response latencies, computed separately for each cell (N = 196 cells; see Supplementary Methods for details of analysis).

Supplementary Figure 9, Salzman



Supplementary Figure 9. Image Identity Coding Index plotted against Image Value-Identity Interaction Coding Index for the visual stimulus (**A**) and trace (**B**) intervals for the same 2-way ANOVA shown in Fig. 3A,B. Blue dots, $p < 0.05$, image value-identity interaction coding index. Green dots, $p < 0.05$, image identity coding index. Red dots, $p < 0.05$ for both factors. Yellow dots, n.s. The index values correspond to the percentage of variance accounted for by image identity and by the interaction between image identity and value, respectively.

Supplementary Notes

Supplementary Note 1. We refer to changes in activity related to image value reversals as "value coding" because they occur as monkeys learn the association between a CS and a positive or negative US. Since the USs we employ have either a positive reinforcing value (eliciting licking, an approach behavior) or negative reinforcing value (eliciting blinking, a defensive behavior), changes in CS-related activity upon learning an association reflects the fact that the CS has acquired predictive value to a monkey. If a cell changed activity in the same direction for both images upon image value reversal, we did not classify it as a value-coding cell. Such responses could be related to other processes important for learning. For example, activity that changes in the same direction may reflect arousal or attention, which could be elicited by a change in task contingencies.

Supplementary Note 2. We also characterized trace interval responses on non-reinforced trials compared to responses on negative and positive trials (see Supplementary Figure 2). Although many amygdala neurons differentiated between these conditions, in general, responses to non-reinforced images were more similar to responses to negative images than to positive ones. This finding is consistent with the notion that non-reinforced trials are mildly aversive, since monkeys are normally motivated by rewards. Indeed, when visual stimuli are presented for longer intervals, in both our paradigm and in other studies, monkeys try to avoid trials rewarded less[1] or not at all[2] by either breaking fixation or failing to complete trials correctly (see Supplementary Figure 7).

Supplementary Note 3. Although functional neuroimaging in humans has illuminated aspects of how the amygdala processes visual stimuli in relation to value, the limited spatial and temporal resolution of fMRI cannot elucidate whether individual amygdala neurons respond differentially to positive and negative visual CSs, as we have done here.

Supplementary Methods.

General Methods.

During experiments, monkeys sat in a Plexiglas primate chair (Crist Instruments) with their eyes 57 cm in front of a 21" Sony CRT monitor. Monkeys were under constant visual observation by the experimenter by way of an infrared video camera connected to a monitor that displayed the monkey to the experimenter.

Electrophysiological Recording and Experimental Control

We used the TEMPO (Reflective Computing) package of software running on Dell Optiplex 260 PCs for experimental control, and the Plexon system for neurophysiological recording, signal amplification, filtering, digitizing of spike waveforms, and data collection. The TEMPO software program directed the sequence of events in a trial, governed stimulus presentation, and enforced the behavioral demands of the task. Analog signals, such as those representing eye positions, were directed to A/D boards in the TEMPO system for behavioral control, and in the Plexon system for subsequent data storage and analysis. The TEMPO system sent "event codes" in real time to the Plexon system through a digital I/O interface, so that the Plexon data files contained all data from an experiment.

Visual stimulus presentation was accomplished with the Videosync program available with TEMPO that runs on a dedicated slave PC. Through a digital I/O interface, the TEMPO server PC directs the Videosync program as to what stimulus to deliver and when to display it. The Videosync computer sent a TTL pulse to the TEMPO computer at the beginning of a vertical retrace in which the display changed. Visual latency analyses occurred relative to this TTL pulse. During periods of fixation, the monkey was required to maintain a position of gaze within 3.5 degrees of the fixation spot, as measured with an infrared eye tracker (ASL, Applied Science Laboratories) that captured pupil images at 240 Hz. Images typically occupied an 8 degree square centred over the fovea.

We measured licking by placing the liquid reward tube just away (~1-2 cm) from the monkey's mouth. Every ms, we measured whether the monkey's tongue interrupted an infrared beam of light that passed between the monkey's mouth and the lick tube. During blinks, the eye tracker would lose its signal and output a characteristic voltage, which we identified and analyzed in Matlab. We verified that these signals were blinks by comparing the signals to a video camera feed showing the monkey blinking during experiments.

The location of the amygdala was ascertained in stereotactic coordinates using MRI imaging with each monkey anaesthetized with isoflurane, intubated and monitored in a 1.5 Tesla research magnet in the Columbia University Department of Radiology. We used these images to guide our subsequent placement of recording chambers.

After a 1-2 week postoperative recovery, we returned to the MRI scanner, verified that our chamber was placed over the amygdala, and placed an electrode in the brain dorsal to the amygdala. To position the electrode, we advanced it through a guide tube that rested in a Crist grid and transversed dura (Crist Instruments). The Crist grid has holes in a grid formation separated by 1 mm. We placed the electrode a known distance

into the brain and acquired images which visualized the electrode in the brain directed at the amygdala. We used two different sequences to visualize the amygdala and tungsten electrodes. A 2D SPGR sequence, with 0.7mm slice thickness, 0 intergap spacing, and 0.234 mm X 0.234 mm within slice pixels, easily revealed the amygdala and a large susceptibility artefact from an electrode (e.g. Fig. 1B). A 2D IR sequence, with 2mm thick slices and 0 intergap spacing and 0.234 mm x 0.468 mm within-slice resolution, still visualized electrodes (e.g. Fig. 1C), but also revealed some anatomical structure within the amygdala (e.g. Fig. 1D). A fluid-filled Crist grid was visible at the top of images using both sequences (see Figs. 1C,D). From the MRIs we obtained, we calculated the distance that the electrode must travel from the end of the Crist grid and the end of the guide tube to the amygdala, and we also noted grey and white matter transitions to guide electrode positioning. So that we could position the electrode in the brain at coordinate locations as determined in the MRI scanner, we also used a grid that supported guide tubes during recording sessions in the lab. We employed either the Crist grid (when recording with one electrode), or a grid supplied with the NAN microdrive in which grid holes were spaced 1.3 mm apart. 4 electrodes separated by 0.34 mm were independently advanced through a single guide tube in the NAN grid.

Data Analysis

Visual Response Latency. To estimate visual response latency, we compared activity during the 500 ms preceding visual stimulus onset to activity during the presentation of visual stimuli. First, for the activity preceding fixation point onset, we constructed a histogram of the number of spikes in each 20 ms bin, shifted by 1 ms across the entire 500 ms period. We then defined a criterion response as one that exceeded 99% of the bins in the histogram (for excitatory responses) or that was less than 95% of the bins (for inhibitory responses). Next, we determined which 20 ms bins, slid in 1 ms steps, met the criterion response during the time from visual stimulus onset until stimulus offset. We defined latency as occurring at the beginning of the first of 20 consecutive overlapping bins that met a criterion response. We computed the latency separately for each image for all cells we recorded. If a cell responded to more than one image, we took the average latency as the estimate of response latency for that cell. We selected 90 ms after visual stimulus appearance as the beginning of the visual stimulus interval because ~95% of the latencies were greater than 90 ms (Supplementary Figure 8). This approach ensured that we would analyze the exact same time intervals in every experiment.

ANOVA Analysis. To determine the relative contribution of image identity and image value to neuronal responses, we performed a 2-way ANOVA on every cell, analyzing the responses to both images that reversed in value. For each cell, this analysis compared activity from the 20 trials before and after the identified change point for each image. If the change point test detected a significant change in activity for only one image, then for the other image we examined activity from the 20 trials before and after the change point identified for the first image. If no change points were detected for either image, then we compared activity from the 20 trials before and after reversal for both images.

ROC Analysis. To determine whether a given cell responded more strongly when an image had a negative or positive value, we performed an ROC analysis that compared activity from the 20 trials before and after an identified change point for a given image. We always designated the negative value trials as the reference distribution. Consequently, ROC values > 0.5 indicated that a cell responded more when an image was positive than when the same image was negative. By contrast, ROC values < 0.5 corresponded to cells that responded more strongly when an image was negative compared to positive. 5 cells changed activity in opposite directions in the visual stimulus and trace interval (ROC values significantly different from 0.5), and we excluded those cells from all analyses.

Permutation Test on Computed ROC Values. We used a permutation test to evaluate whether each ROC value was significantly different from 0.5 (Fig. 3C). For this analysis, we abolished the relationship between neural activity and image value by randomly assigning each neural response a value, but keeping constant the overall distribution of neural activity and the number of responses assigned to each value. We then computed the ROC value for this permuted data set, and repeated the procedure 1000 times. An ROC value on the original data set that fell outside the central 95% of the distribution of ROCs computed on the permuted data was considered statistically significant.

Normalization Procedure. When comparing activity across the population of neurons, we normalized activity for each cell by first dividing each trial's activity by the median response across trials, and subsequently subtracting the mean of the normalized values across all trials from each trial's normalized value. For all images, we analyzed activity from the 20 trials before and after the reversal in image value. Note that we also multiplied by -1 the normalized activity from the neurons responding more strongly before reversal on both reinforced and non-reinforced trials (so that activity went from low-to-high for all cells). Since all trial types were randomly interleaved, the change in activity on positive and negative images could not be due to changes in overall cell responsivity as a result of the reversal, or we would have observed similar changes on the non-reinforced images. We performed a similar procedure on the behavioural data, except normalizing by the mean of the data values. We fit Weibull functions to the behavioural and neural data, with four free parameters (see equation 1). l and u parameters adjusted the lower and upper asymptotes, and α and β influenced the shape of the function (α adjusted the latency of the rise of the function, and β the abruptness of the onset of the rise). Population learning curves for licking and blinking were extremely similar, so they were combined for this analysis (95% prediction intervals of the two curves overlapped at every trial). Similarly, the 95% prediction intervals for curves describing neural activity and behaviour overlapped at every trial for visual stimulus and trace interval activity both for images that changed from positive-to-negative and for images that changed from negative-to-positive. Consequently, we combined all these data to yield one neural learning curve and one behavioural learning curve (Fig. 4C).

References

1. Roesch, M. R. & Olson, C. R. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304**, 307-10 (2004).
2. Shidara, M., Aigner, T. G. & Richmond, B. J. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *Journal of Neuroscience* **18**, 2613-25 (1998).