ARTICLE

doi:10.1038/nature21682

Neural ensemble dynamics underlying a long-term associative memory

Benjamin F. Grewe[1,2,3], Jan Gründemann[4], Lacey J. Kitch[1,2,3], Jerome A. Lecoq[1,2,3], Jones G. Parker[3,5], Jesse D. Marshall[1,2,3], Margaret C. Larkin[1,3], Pablo E. Jercog[1,2,3], Francois Grenier[4]†, Jin Zhong Li[1,3], Andreas Lüthi[4,6] & Mark J. Schnitzer[1,2,3]

The brain's ability to associate different stimuli is vital for long-term memory, but how neural ensembles encode associative memories is unknown. Here we studied how cell ensembles in the basal and lateral amygdala encode associations between conditioned and unconditioned stimuli (CS and US, respectively). Using a miniature fluorescence microscope, we tracked the Ca^{2+} dynamics of ensembles of amygdalar neurons during fear learning and extinction over 6 days in behaving mice. Fear conditioning induced both up- and down-regulation of individual cells' CS-evoked responses. This bi-directional plasticity mainly occurred after conditioning, and reshaped the neural ensemble representation of the CS to become more similar to the US representation. During extinction training with repetitive CS presentations, the CS representation became more distinctive without reverting to its original form. Throughout the experiments, the strength of the ensemble-encoded CS–US association predicted the level of behavioural conditioning in each mouse. These findings support a supervised learning model in which activation of the US representation guides the transformation of the CS representation.

Associative fear conditioning induces a long-term memory that requires the basal and lateral amygdala (BLA)[1–3] but not hippocampal[4] activity. Previous studies found BLA neurons with potentiated responses to a CS, such as an auditory tone, after associative conditioning with an aversive US[1–3]. This prompted a Hebbian model in which 'fear cells' with co-active inputs conveying the paired CS–US presentations potentiate their responses to subsequent CS presentations[1,3,5]. However, the dynamics of individual fear cells seem too stochastic to support reliable memory storage[1]. Neural ensembles might allow more robust storage, but how cell ensembles encode associative memories and whether this fits the Hebbian model remain unknown.

To track BLA neural ensemble activity in behaving mice, we combined time-lapse microendoscopy, a head-mounted microscope[6,7] and expression of the GCaMP6m Ca^{2+} indicator[8] in excitatory neurons (Fig. 1a, b; Extended Data Fig. 1; Methods). This differs from previous electrophysiological studies of BLA that lacked access to ensemble activity patterns and had limited recording durations[1], and from studies of immediate early gene activation[9,10], which poorly reports declines, temporal patterns and gradations of electrical activity.

We first examined neural responses to tones and electric shocks in awake mice (Extended Data Fig. 2). The cells that responded to these stimuli were sparse and interspersed across the BLA[10,11]. This intermingling may help the BLA to link temporally associated signals of different types via local circuit interactions[10–14].

To study associative memory[1,12,14–16], we repeatedly paired an auditory cue (CS^+; 25 × 200-ms-tone-pulses per presentation) with a footshock US. As a control, we repeatedly presented another tone (CS^-) without the US (Fig. 1c). Mice with and without implanted microendoscopes had comparable expression of CS^+-evoked fear responses, visible as conditioned freezing[15,17] (Extended Data Fig. 3). Across a 6-day protocol, cells responding to the CS^+ or CS^- ($P \leq 0.01$, evoked signals versus baseline, rank-sum test) were sparse, interspersed and largely distinct (Fig. 1c–e). CS-evoked Ca^{2+} transients closely resembled those expected from previous electrical recordings[12] (Extended Data Fig. 4).

Across all 6 days, the number of active cells remained constant (152 ± 14 cells per day per mouse (mean ± s.e.m.); Friedman test; 12 mice; see Supplementary Table 1 for χ^2 and P values), but after conditioning approximately 45% more cells responded[1,10] to the CS^+ (Fig. 2) (before training, $9 \pm 1\%$ cells were CS^+-responsive versus $14 \pm 1\%$ afterwards; $P \leq 0.01$, rank-sum test; 2 pre- and 3 post-training sessions). The percentages of CS^--responsive cells also rose, paralleling the small increase in CS^--evoked freezing above baseline levels and suggesting that the CS^- was not a learned safety signal[18] (Figs 1c, 2a; Extended Data Fig. 3g; Supplementary Note). During conditioning (day 3), $14 \pm 3\%$ of active cells responded to the US; within this subset a minority up- ($7 \pm 3\%$) or down-regulated ($13 \pm 5\%$) these responses during training (Fig. 2c, d).

Using image alignment, we registered cell identities over the 6 days (171–438 cells per mouse; 3,655 total; 12 mice). Similar percentages of cells were active each day ($49 \pm 2\%$; Extended Data Fig. 5; Supplementary Table 1). A multitude of cells was active on 1–2 days ($49 \pm 3\%$), and a minority on all days ($16 \pm 2\%$). Individual cells came in and out of the active ensemble day-to-day; there were about 55% cells in common for consecutive sessions and 35% for 5 days apart. This turnover resembles that seen in long-term studies of the hippocampus[7,19] and might be a general phenomenon in brain areas processing long-term memories.

We next studied the encoding of the CS^+–US association and tested the Hebbian model[20]. Notably, only $38 \pm 5\%$ of cells with heightened CS^+-evoked responses after training responded to the US during training, whereas $65 \pm 6\%$ of cells that were initially responsive to both the CS^+ and US were less CS^+-responsive after training (Extended Data Fig. 6). Of cells with significant responses to the CS^+ on at least one day, $32 \pm 2\%$ potentiated these responses after training, whereas $28 \pm 7\%$ reduced them (Fig. 2d; $P \leq 0.05$, rank-sum test; 125 CS^+ tone-pulses per day before training, 300 afterwards). CS^--responsive cells underwent analogous changes, to a lesser extent (Fig. 2d). Overall, this bi-directional plasticity was unpredicted from Hebbian potentiation[20].

[1]James H. Clark Center for Biomedical Engineering & Sciences, Stanford University, Stanford, California, USA. [2]Howard Hughes Medical Institute, Stanford University, Stanford, California, USA. [3]CNC Program, Stanford University, Stanford, California, USA. [4]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [5]Pfizer Neuroscience Research, Cambridge, Massachusetts, USA. [6]University of Basel, Basel, Switzerland. †Present address: International Institute for Integrative Sleep Medicine (WPI-IIIS), University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Japan.

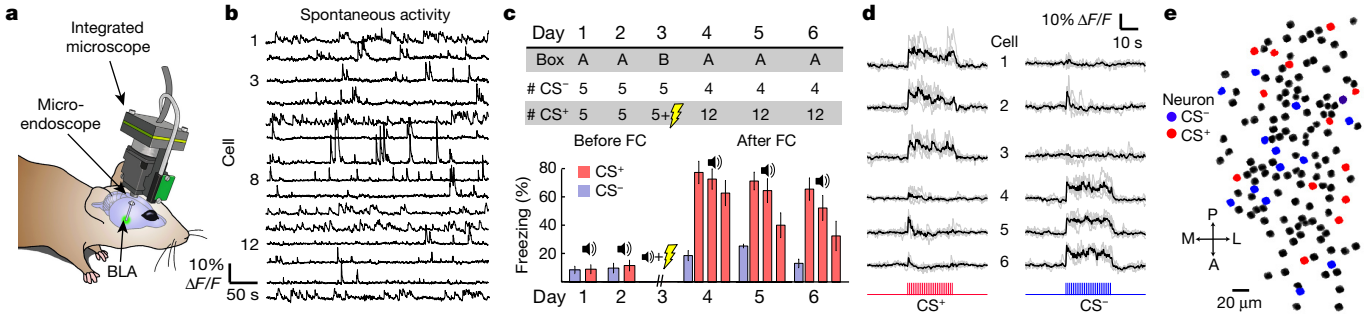


Figure 1 | Ca^{2+} imaging of BLA neural activity across a 6-day fear-conditioning protocol. a, A miniature microscope and implanted microendoscope allowed large-scale neural Ca^{2+} imaging. **b**, Traces of spontaneous Ca^{2+} activity from 15 BLA neurons. **c**, Top, fear conditioning (FC) protocol, with numbers of stimuli shown. Bottom, percentages of time 12 mice froze during CS^+ and CS^- presentations. The occurrences of the auditory CS and the foot-shock US are denoted by 'sound' and 'lightning' symbols, respectively. Values are respectively averaged over 5 and 4 stimulus presentations, before and after conditioning; data are mean ± s.e.m. **d**, Activity traces of cells responsive to CS^+ or CS^- presentations before conditioning. **e**, Map of BLA cells in one mouse. Coloured cells responded to CS^+ or CS^- tones. A, anterior; L, lateral; M, medial; P, posterior. Traces in **b** and **d** were downsampled to 200-ms time bins.

To study ensemble coding, we tested whether CS^+ and CS^- presentations were identifiable from their evoked activity patterns. We trained three-way, Fisher linear decoders to distinguish baseline conditions from CS^+ and CS^- presentations on each day. These decoders classified the three conditions accurately (97 ± 0.3% of 1-s time segments) for all 6 days (Fig. 3a). Accuracy fell slightly using only CS^+- and CS^--responsive cells (90 ± 3% accuracy), but substantially when we omitted all CS-responsive cells (61 ± 2% accuracy). Across the first five tone-pulses of each CS presentation, decoding accuracy and conditioned freezing both rose to an asymptote (Extended Data Fig. 7), suggesting that BLA coding fidelity improved as the tones were repeated within a CS presentation.

How did conditioning affect ensemble coding? To investigate separately CS^+ and CS^- encoding, we trained two sets of binary decoders, which discriminated either CS^+ or CS^- presentations from baseline conditions. We trained each decoder on data from one day and tested it on data from other days. Despite day-to-day fluctuations in the cells, CS^- decoders had up to 85% accuracy across days (74 ± 1%; Extended Data Fig. 8). CS^+ decoders performed similarly, provided that the training and testing days were both before or both after conditioning (74 ± 1% accuracy), but if they spanned the conditioning session, accuracy fell to chance levels (55 ± 1%) (Fig. 3b). Hence, representations of the CS^+, but not the CS^-, changed significantly during memory formation, consistent with the bi-directional plasticity of CS^+-responsive cells.

To study plasticity further, we constructed multi-dimensional population vectors (one dimension per cell) for each response to a CS or US. To assess the differentiability of the responses, we used the Mahalanobis population vector distance (PVD)[21]. This resembles an Euclidean distance, but like the discriminability index (d') from statistics it accounts for mean differences and trial-to-trial variability[21], using the correlations in the responses of the cells (Extended Data Figs 8, 9). To examine how training changed the CS^+ representation, we divided day 3 into early and late training phases and computed the mean PVDs between US- and CS^+-evoked responses in each phase (Fig. 3c). Notably, training increased the similarity and decreased the discriminability of the US and CS^+ representations. Across five CS^+–US pairings, PVDs declined by a significant amount (Δ_1: −8 ± 2%; $P = 0.02$, signed-rank test; 12 mice, early versus late mean PVDs; 3,655 cells), owing to increased similarity of the mean responses to CS^+ and US, not due to decreases in their variability (Extended Data Fig. 9). CS^- representations remained invariant ($\Delta_1 = -0.2 \pm 0.3\%$; $P = 0.3$).

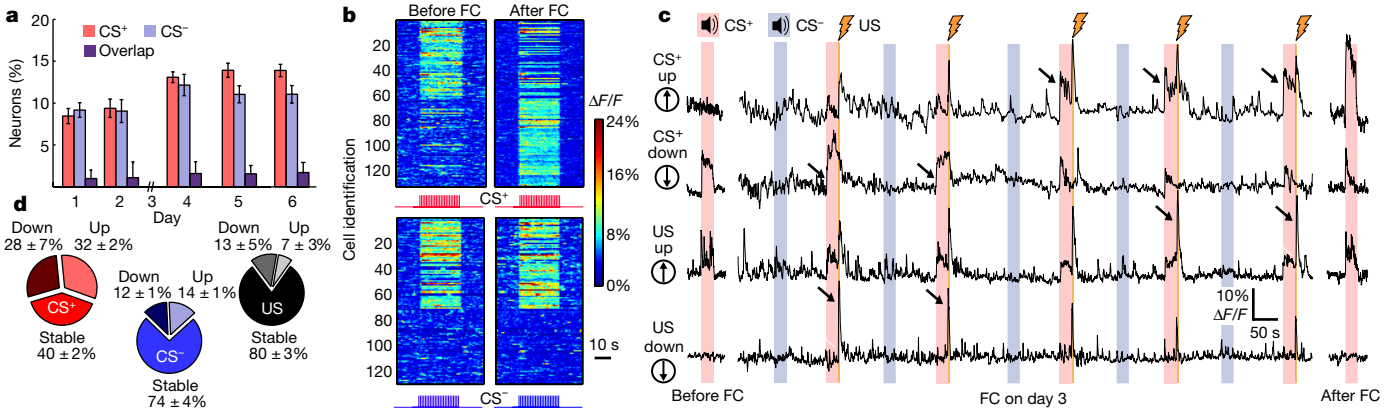


Figure 2 | Fear conditioning induces bi-directional changes in BLA signalling. a, Percentages of cells responding to the CS^+, CS^- or both stimuli. **b**, Ca^{2+} signals, showing changes in CS^+ encoding and stable CS^- encoding for 2 sets of 125 cells detected throughout the study. Top, cells responsive to the CS^+ on at least one day. Bottom, cells that either responded to the CS^- on one or more days, or lacked responses to both CS types. Colours show the Ca^{2+} response of each cell averaged over 5 CS presentations on the day the cell responded maximally, for days before and after fear conditioning. Cells are arranged by whether they responded maximally before or after conditioning. **c**, Ca^{2+} signals from four cells, before (left, mean over 5 CS^+ presentations), during (middle, single trial), and after (right, mean over 5 CS^+ presentations) conditioning, illustrating altered responses to the CS^+ (top two traces) or US (bottom two traces). **d**, Percentages of cells after conditioning with stable, increased or decreased responses to the CS^+ (red), CS^- (blue) and US (black), respectively, based on 231, 362 and 261 neurons. Cells in the former two charts responded to the CS on at least one day before or after conditioning. Cells in the latter chart responded significantly to the US on day 3. Traces in **b** and **c** were downsampled to 200-ms time bins. **a**, **b** and **d** are from $n = 12$ mice. Data in **a** and **d** are mean ± s.e.m.

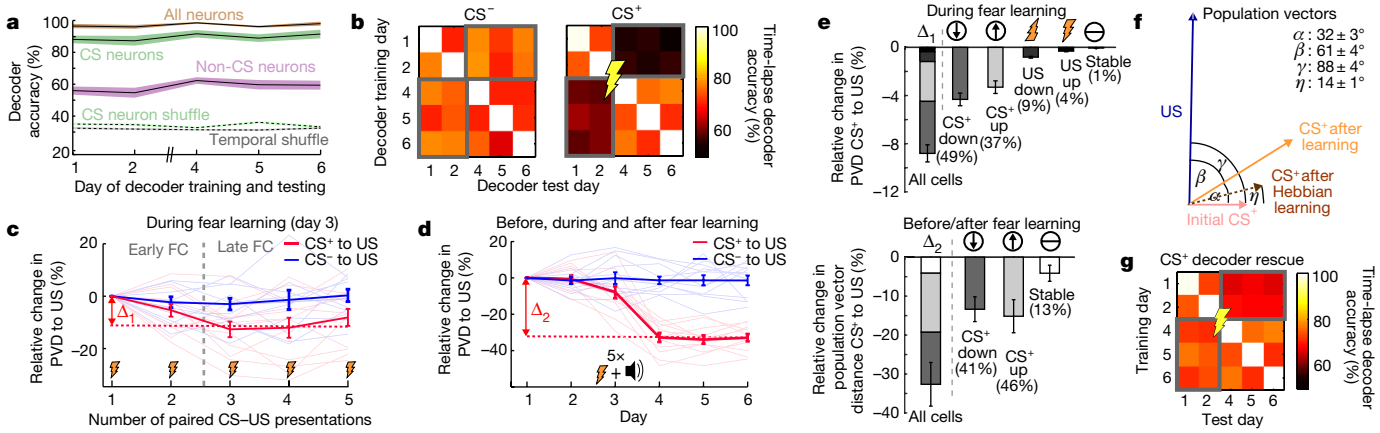


Figure 3 | Learning increases the similarity of the CS$^+$ and US representations. a, Accuracies of same-day, three-way decoders discriminating baseline, CS$^+$ and CS$^-$ presentations. Decoders based on CS-responsive cells (green curve) nearly matched those using all cells (brown). Decoders based on cells unresponsive to the CS (purple) performed poorly, but better than decoders given temporally shuffled Ca^{2+} traces (grey dotted line) or shuffled cell identities (green dotted line). 152 ± 14 cells per day per mouse (3,655 cells total; 12 mice). Shading denotes s.e.m. **b**, Accuracies of inter-day, binary decoders distinguishing CS$^-$ (left) or CS$^+$ (right) presentations from baseline conditions. **c**, Population vector distances (PVDs) between US-evoked ensemble activity and that evoked by the CS$^-$ (blue) or CS$^+$ (red) during conditioning (day 3). CS$^+$–US PVDs declined by an amount Δ_1 as responses to the two stimuli became more similar. Dashed vertical line separates early and late CS$^+$–US pairings; to calculate Δ_1, we compared these two portions of the session. 155 ± 11 cells per mouse on day 3 (1,860 cells total; 12 mice). **d**, CS$^+$–US PVDs declined during and after training, indicating increased similarity of the two representations. Δ_2 is the difference in PVD values before versus after training (3,655 cells). **e**, Top, composition of the changes, Δ_1, in CS$^+$–US PVDs between early and late phases of training, defined in **c**. Bottom, analogous graph for Δ_2, showing how CS$^+$–US PVDs changed from before (days 1, 2) to after (days 4–6) conditioning. To decompose Δ_1 and Δ_2, we examined cells with stable (white), up- (light grey) or downregulated (dark grey) responses to the CS$^+$, and cells with up- or downregulated responses to the US (black). Error bars in **c**–**e** denote s.e.m. **f**, Before conditioning, population vector representations of the US (blue) and CS$^+$ (pink) were orthogonal (88° ± 4°; 12 mice). Afterwards, the CS$^+$ population vector (orange) was 210 ± 20% longer, rotated 32° ± 3° from its initial orientation, and had a 61° ± 4° angle to the US representation, indicating the rotation was in the plane defined by the US representation and that of the initial CS$^+$. These changes differed from predictions of Hebbian potentiation (maroon) (angle and length changes are all $P < 10^{-4}$, rank-sum test). **g**, Mean accuracies of time-lapse decoders after computational rescue of their ability to distinguish CS$^+$ presentations from baseline. For each pairing of one pre- and one post-training day (pairs inside grey rectangles), we rescued population vectors from the testing day by applying the optimal transformation, determined as in Extended Data Fig. 10c.

Even larger coding changes occurred after training. By day 4, CS$^+$ and US representations were 32% less differentiable than before training (Fig. 3d, e), unforeseen from studies of consolidation that suggested a stabilization of neural coding[22]. Of the total change in CS$^+$–US PVD (Δ_2), 75% first appeared on day 4 (Fig. 3d) (CS$^+$: $\Delta_2 = -32 \pm 6\%$ relative to day 1 PVDs; CS$^-$: $\Delta_2 = 0.5 \pm 5\%$; 2 PVDs before training and 3 afterwards, in each of 12 mice). On days 4–6, the CS$^+$ population vector had increased 210 ± 20% (12 mice) in amplitude and rotated (32° ± 3°) nearly directly towards the US population vector (Fig. 3f). The re-scaling reflected increased CS$^+$-evoked responses of many cells that never responded to the US, tempered by the decreased CS$^+$-evoked responses of other cells. The rotation towards the US representation reflected new CS$^+$-evoked responses in cells previously lacking them. These changes differed from the predictions of Hebbian potentiation (changes in vector length and angle, each $P < 10^{-4}$, rank-sum test; 12 mice).

Cells with decreased CS$^+$-evoked responses and cells with increased CS$^+$-evoked responses were equally important for the re-coding (Fig. 3e), during training ($P = 0.2$, signed-rank test, comparing contributions to Δ_1 of cells with up- (37 ± 2%) versus downregulated (49 ± 2%) CS$^+$ responses), and during consolidation of learning ($P = 0.9$, contributions to Δ_2 of cells with up- (46 ± 2%) versus downregulated (41 ± 2%) CS$^+$ responses). Changes in US encoding made smaller (13 ± 1%) but still significant ($P = 0.008$) contributions to the similarity increase between CS$^+$ and US representations.

We next investigated how the CS$^+$ encoding changes that occurred during learning consolidation related to those from training. We hypothesized that consolidation proportionally accentuates changes from training. To test this, we linearly extrapolated the changes to the CS$^+$ representation from conditioning (Δ_A) and examined how well this captured the consolidated responses (Extended Data Fig. 10). Successful extrapolations should rescue the unsuccessful time-lapse CS$^+$ decoders trained and tested on days spanning conditioning. With extrapolations 4–5 times Δ_A in amplitude, CS$^+$ decoding reached 72 ± 3% accuracy, nearing that of time-lapse CS$^-$ decoders (74 ± 1%) (Fig. 3g). If we limited extrapolation to cells with only up- or only downregulated CS$^+$-evoked responses, the rescue of CS$^+$ decoding badly degraded, highlighting the importance of bi-directional plasticity during consolidation.

On days 4–6, mice underwent partial behavioural extinction, comprising acute (within session) and consolidated (across session) effects (Figs 1c, 4, 5). Did this reflect a change in the encoded CS$^+$–US association? As found previously[1], individual cells up- or downregulated their CS$^+$-evoked responses during acute extinction (Fig. 4a). We assessed how this affected CS–US PVDs across 4 CS$^-$ and 12 unreinforced CS$^+$ presentations. Between the first four and last four CS$^+$ presentations, the CS$^+$ and US representations became significantly more differentiable ($\Delta_3 = 20 \pm 1\%$, normalized to the mean PVD on day 1; $P < 10^{-3}$, signed-rank test; 144 early versus 144 late CS$^+$ presentations on days 4–6; Fig. 4b). This acute change reflected an 18 ± 5% (12 mice) reduction in CS$^+$ population vector amplitude and an 8° ± 3° rotation away from the US vector. These changes were absent for the CS$^-$ ($\Delta_3 = -3 \pm 3\%$; $P = 0.3$).

Unlike acute fear learning, during acute extinction cells with decreased CS$^+$-evoked responses contributed more to the CS$^+$ representation changes than cells with increased responses (Fig. 4c). However, the rates at which ensemble coding changed during acute learning (day 3) and extinction were equivalent ($P = 0.6$; Fig. 4d), suggesting a common process for initial storage of a memory and its acute extinction. During within-session extinction, the CS$^+$ representation did not revert and gained no more similarity to its initial representation before learning (Fig. 4e) ($\Delta_4 = -2 \pm 2\%$; $P = 0.37$, Friedman test). Instead, the CS$^+$ population vector rotated out of the plane defined by

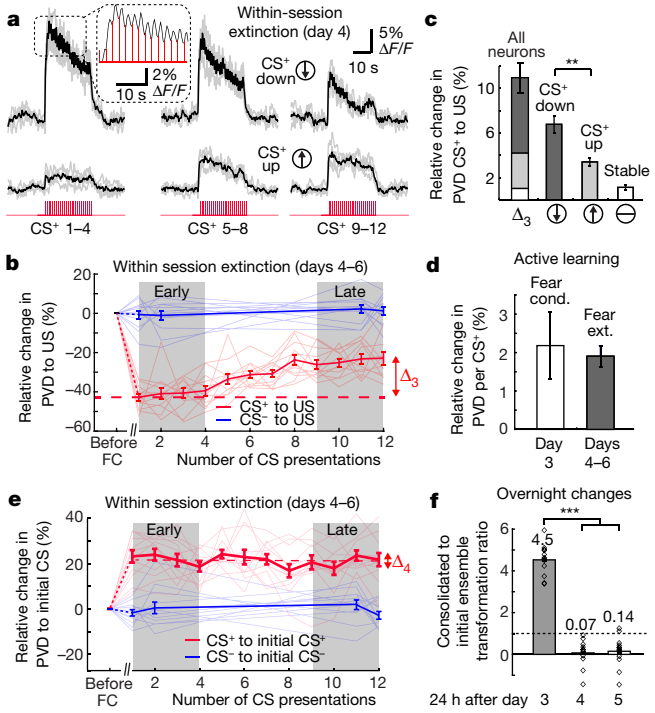


Figure 4 | During behavioural extinction, the CS$^+$ representation becomes more distinguishable from the US representation but does not revert to its initial form. a, Ca^{2+} signals from two neurons, illustrating bi-directional plasticity of CS$^+$-evoked responses in early (left), middle (centre), and late (right) phases of fear extinction on day 4. Grey lines, individual CS$^+$ presentations (4 per set). Black lines, mean responses. Inset, magnified view of responses to individual CS$^+$ tone pulses. **b**, PVDs between CS- and US-evoked activity during extinction (days 4–6), for individual mice (thin lines) and averaged across 12 mice (thick lines) for 12 CS$^+$ and 4 CS$^-$ presentations per day (2,181 total cells). CS$^+$–US PVDs (red lines) increased by an amount (Δ_3) within the sessions after fear conditioning. CS$^-$–US PVDs (blue lines) were stable. **c**, Composition of the change in PVDs (Δ_3) in **b**, from cells with stable (white), increased (light grey), or decreased (dark grey) responses to the CS$^+$ after training. $^{**}P < 0.001$, signed-rank test ($n = 12$ mice). **d**, Within individual sessions, the CS$^+$ representation changed at similar rates during learning and extinction (quantified by the change in mean CS$^+$–US PVD per CS$^+$ presentation; $P = 0.6$, signed-rank test; 12 mice). **e**, During extinction sessions (days 4–6), there was little change (Δ_4) in the mean PVDs (thick lines) between CS$^+$ and CS$^-$ representations and their initial forms before conditioning (averaged over days 1 and 2). Thin lines denote data from individual mice. **f**, Overnight consolidation induced long-term changes in the CS$^+$ representation 24 h after conditioning (day 3) but not after extinction training (days 4 and 5). Horizontal line marks where coding changes from training are neither amplified nor diminished in consolidation. Numbers above each dataset denote mean coding changes occurring overnight after each day, that is, a 450% increase after day 3, and reductions to 7% and 14% of their values after Ca^{2+} imaging on days 4 and 5. Open diamonds: values from 12 individual mice. $^{***}P < 0.001$, signed-rank test, day 3 versus days 4 or 5. All error bars denote s.e.m.

the US and the initial CS$^+$ (Fig. 5e), maintaining a $28° \pm 3°$ angle to its initial form that differed little from the $32° \pm 3°$ at the end of learning.

Hence, ensembles of BLA neurons explicitly encode extinction training as new learning[1]. We did not find overt signals of US omission, but sub-threshold signals might drive plasticity in an extinction-specific subset of cells[1] (Fig. 4a). Extinction engages the hippocampus, thalamus and neocortex[23], and their inputs to the BLA might signal US omission. Unlike consolidation of learning, most coding changes that accumulated in each extinction session reversed before the next session (Fig. 4f), consistent with the modest behavioural extinction that persisted overnight (Figs 1c, 5d).

We examined how encoding of the CS$^+$–US association related to conditioned behaviour. The differentiability of the two representations predicted the overall extent of freezing behaviour, throughout learning and extinction ($r = 0.7$; $P < 10^{-14}$; Fig. 5a). Yet, on a time scale of seconds, the mean CS$^+$–US PVD values were no different between freezing and non-freezing epochs (Fig. 5b). Thus, resemblance of the CS$^+$- and US-encodings predicts the general acquisition, not the instantaneous performance, of learned freezing[24]. How much the CS$^+$-encoding altered its similarity to the US-encoding strongly predicted the behaviour of individual mice during learning and extinction (Fig. 5c, d).

Discussion

On the basis of recordings of more than 3,600 BLA cells across 6 days, the analyses here show how ensembles of neurons represent associative information. The sets of active and CS$^-$-responsive neurons exhibited day-to-day turnover, but the neural ensembles encoded information far more reliably than individual cells[7,19,25,26]. It is unclear what mechanisms preserve information despite cellular turnover, which might reflect variations in immediate early gene expression that help time-stamp individual memories[26–29].

Single-cell recordings have shown that neurons in several amygdalar areas can individually depress or potentiate their response properties under various conditions, leading to the impression that depression and potentiation may result from opposing influences on memory storage[1,30,31]. The recordings here show that learning simultaneously induces potentiation and depression of CS$^+$-evoked responses of cells in an equally balanced manner (Figs 2a, 3e). This coordinated bi-directional plasticity was crucial for transforming the ensemble level CS$^+$ representation to increase its similarity to the US representation (Fig. 3f), was undetectable in previous studies using immediate early gene activation[10] or pharmacological inactivation methods[15,17], and mainly occurred during consolidation of learning (Fig. 3d–g).

Notably, our results diverge from the predictions of Hebbian fear-learning[1,2,27,32], which invokes a bi-conditional rule requiring coincident CS$^+$ and US signals and posits that among cells receiving CS$^+$ signals, those activated by the US will potentiate their CS$^+$-evoked responses[20]. Mechanisms associated with this rule, such as NMDA-receptor-dependent synaptic potentiation[32], probably contribute to transforming the CS$^+$ representation, but the basic Hebb rule alone does not predict all the observed plasticity.

First, up- and downregulation of stimulus-evoked responses were equally prevalent and important for transforming coding during learning (Fig. 3e). Second, most cells with potentiated CS$^+$ responses were unresponsive to the US (Fig. 3f; Extended Data Fig. 6). Third, a majority of cells that were CS$^+$- and US-responsive before training had reduced CS$^+$-evoked responses afterwards. Fourth, bi-conditional plasticity rules have difficulty explaining why many CS$^+$-responsive cells depress their responses but CS$^-$-responsive cells generally do not (Fig. 2a, d). A mere lack of US-related input cannot explain this difference. Hebbian models require coincident CS$^+$ and US inputs to induce potentiation[20], but, in reality, amygdala-dependent fear-learning does not require coincidence[4,33]. Explaining this temporal permissiveness and the differences in plasticity between CS$^+$- and CS$^-$-responsive cells probably requires a modified Hebb rule.

One possibility is a tri-conditional rule that refers not only to CS and US presentations but also to a third factor, such as a neuromodulator or network-wide US-evoked inhibition[14,34–36], to explain the plasticity differences between CS$^+$- and CS$^-$-responsive cells (Supplementary Table 2; Supplementary Note). Theorists have studied such 'neo-Hebbian' tri-conditional rules[37], and both inhibitory signalling and neuromodulator release are crucial for fear-learning-induced changes to occur in the BLA at normal rates[34,35]. Our data suggest that these network-wide factors might aid ensemble encoding by promoting bi-directional plasticity for CS$^+$–US pairings in close but not strict

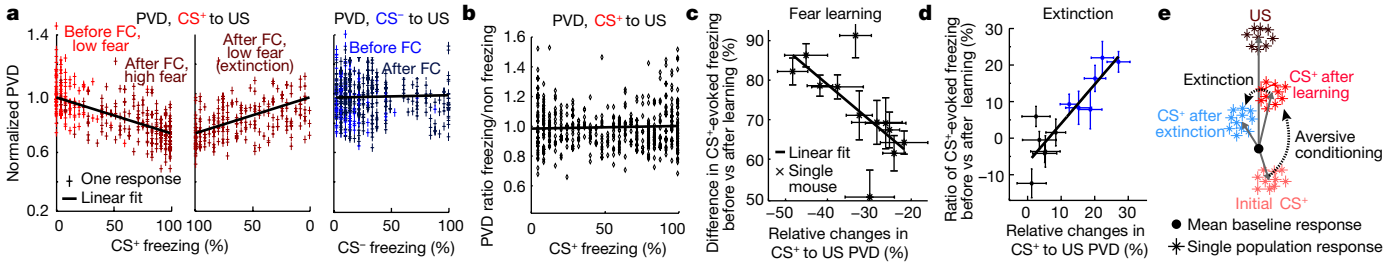


Figure 5 | The similarity of the CS$^+$ and US representations encodes the CS$^+$–US association strength. a, CS$^+$–US PVDs for each CS$^+$ presentation, normalized to the CS$^+$–US PVD for the first CS$^+$ presentation of the mouse, are predictive of the freezing level that each CS$^+$ evoked before and during conditioning (left), and during extinction (middle). CS$^-$–US PVDs (right) lack this relationship. **a** and **b** are based on 3,655 neurons in 12 mice. Black lines denote linear fits. **b**, Within each 25-s CS$^+$, the 1-s time bins with and without freezing had a near unity ratio between their CS$^+$–US PVD values, irrespective of the evoked freezing level. **c**, How much each mouse (individual data points) exhibited post-training changes in the CS$^+$–US PVD was predictive of its learned, CS$^+$-evoked freezing behaviour. Black line denotes linear fit ($r = 0.7$; $P < 10^{-3}$). Error bars denote s.e.m. across six pair-wise comparisons of one day before (days 1 and 2) and one day after (days 4–6) training for each mouse. **d**, How much each mouse (data points) exhibited a changed CS$^+$–US PVD between the first six CS$^+$ on day 4 versus the first six CS$^+$ on day 6 was predictive of its loss of CS$^+$-evoked freezing. Black line denotes linear fit ($r = 0.9$; $P < 10^{-3}$). Blue points denote mice with significant consolidated extinction ($P < 0.05$, signed-rank test comparing time spent freezing between the two sets of CS$^+$ presentations). Error bars denote s.e.m. across the six pair-wise comparisons of one CS$^+$ from among the first six presentations on day 4 and the corresponding CS$^+$ from the first six presented on day 6. **e**, Schematic of CS$^+$ population vector changes during learning and extinction. During learning, this vector doubles in length. It also rotates directly towards and becomes less differentiable from the US population vector, supporting a model in which the US representation provides a learning supervision signal. During acute extinction, the CS$^+$ population vector shrinks approximately 20% and rotates 5–8° out of the plane defined by the US and the initial CS$^+$.

concurrence[38]. Nevertheless, different cells might follow different plasticity rules, and some might follow the simple Hebb rule.

The data here naturally suggest an abstract interpretation of how associative information is stored and represented, namely that BLA ensembles of neurons implement a supervised learning algorithm[39] to encode the CS–US association. Previous studies proposed that the US acts as a cellular-level teaching signal[20,40]. Here, the plasticity of single cells was not strictly determined by US-evoked activity. Instead, US-driven activity seemed to provide an ensemble-level supervision signal, guiding rotation of the CS$^+$ population vector directly towards the US representation (Figs 3f, 5e), which would have been unapparent in smaller recordings[1,40]. An attraction of this account is its intrinsic measure of memory strength, the similarity of the US and CS$^+$ representations. Conditioned freezing closely tracked the US–CS$^+$ PVD for each mouse, strongly supporting this interpretation. Principles of supervised learning might apply to brain areas beyond the BLA, and future work should examine whether coding transformations similar to those seen here occur in other limbic regions or areas of neocortex.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 June 2016; accepted 1 February 2017.
Published online 22 March 2017.

1. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454,** 600–606 (2008).
2. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nat. Rev. Neurosci.* **5,** 844–852 (2004).
3. Quirk, G. J., Armony, J. L. & LeDoux, J. E. Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron* **19,** 613–624 (1997).
4. Chowdhury, N., Quinn, J. J. & Fanselow, M. S. Dorsal hippocampus involvement in trace fear conditioning with long, but not short, trace intervals in mice. *Behav. Neurosci.* **119,** 1396–1402 (2005).
5. Quirk, G. J., Repa, C. & LeDoux, J. E. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron* **15,** 1029–1039 (1995).
6. Ghosh, K. K. *et al.* Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8,** 871–878 (2011).
7. Ziv, Y. *et al.* Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16,** 264–266 (2013).
8. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).
9. Garner, A. R. *et al.* Generation of a synthetic memory trace. *Science* **335,** 1513–1516 (2012).
10. Gore, F. *et al.* Neural representations of unconditioned stimuli in basolateral amygdala mediate innate and learned responses. *Cell* **162,** 134–145 (2015).
11. Namburi, P. *et al.* A circuit mechanism for differentiating positive and negative associations. *Nature* **520,** 675–678 (2015).
12. Senn, V. *et al.* Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron* **81,** 428–437 (2014).
13. Romanski, L. M., Clugnet, M. C., Bordi, F. & LeDoux, J. E. Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behav. Neurosci.* **107,** 444–450 (1993).
14. Wolff, S. B. *et al.* Amygdala interneuron subtypes control fear learning through disinhibition. *Nature* **509,** 453–458 (2014).
15. Blair, H. T. *et al.* Unilateral storage of fear memories by the amygdala. *J. Neurosci.* **25,** 4198–4205 (2005).
16. Ciocchi, S. *et al.* Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468,** 277–282 (2010).
17. Erlich, J. C., Bush, D. E. & Ledoux, J. E. The role of the lateral amygdala in the retrieval and maintenance of fear-memories formed by repeated probabilistic reinforcement. *Front. Behav. Neurosci.* **6,** 16 (2012).
18. Kong, E., Monje, F. J., Hirsch, J. & Pollak, D. D. Learning not to fear: neural correlates of learned safety. *Neuropsychopharmacology* **39,** 515–527 (2014).
19. Mankin, E. A. *et al.* Neuronal code for extended time in the hippocampus. *Proc. Natl Acad. Sci. USA* **109,** 19462–19467 (2012).
20. Blair, H. T., Schafe, G. E., Bauer, E. P., Rodrigues, S. M. & LeDoux, J. E. Synaptic plasticity in the lateral amygdala: a cellular hypothesis of fear conditioning. *Learn. Mem.* **8,** 229–242 (2001).
21. Bishop, C. M. *Pattern Recognition and Machine Learning* Vol. 1 (Springer, 2007).
22. Rodrigues, S. M., Schafe, G. E. & LeDoux, J. E. Molecular mechanisms underlying emotional learning and memory in the lateral amygdala. *Neuron* **44,** 75–91 (2004).
23. Sotres-Bayon, F. & Quirk, G. J. Prefrontal control of fear: more than just extinction. *Curr. Opin. Neurobiol.* **20,** 231–235 (2010).
24. Goosens, K. A., Hobin, J. A. & Maren, S. Auditory-evoked spike firing in the lateral amygdala and Pavlovian fear conditioning: mnemonic code or fear bias? *Neuron* **40,** 1013–1022 (2003).
25. Danielson, N. B. *et al.* Sublayer-specific coding dynamics during spatial navigation and learning in hippocampal area CA1. *Neuron* **91,** 652–665 (2016).
26. Rubin, A., Geva, N., Sheintuch, L. & Ziv, Y. Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife* **4,** e12247 (2015).
27. Han, J. H. *et al.* Neuronal competition and selection during memory formation. *Science* **316,** 457–460 (2007).
28. Rashid, A. J. *et al.* Competition between engrams influences fear memory formation and recall. *Science* **353,** 383–387 (2016).
29. Rogerson, T. *et al.* Molecular and cellular mechanisms for trapping and activating emotional memories. *PLoS One* **11,** e0161655 (2016).
30. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439,** 865–870 (2006).
31. Belova, M. A., Paton, J. J., Morrison, S. E. & Salzman, C. D. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* **55,** 970–984 (2007).
32. Rumpel, S., LeDoux, J., Zador, A. & Malinow, R. Postsynaptic receptor trafficking underlying a form of associative learning. *Science* **308,** 83–88 (2005).

33. Sah, P., Westbrook, R. F. & Lüthi, A. Fear conditioning and long-term potentiation in the amygdala: what really is the connection? *Ann. NY Acad. Sci.* **1129,** 88–95 (2008).
34. Fadok, J. P., Dickerson, T. M. & Palmiter, R. D. Dopamine is necessary for cue-dependent fear conditioning. *J. Neurosci.* **29,** 11089–11097 (2009).
35. Johansen, J. P. *et al.* Hebbian and neuromodulatory mechanisms interact to trigger associative memory formation. *Proc. Natl Acad. Sci. USA* **111,** E5584–E5592 (2014).
36. Gütig, R. & Sompolinsky, H. The tempotron: a neuron that learns spike timing-based decisions. *Nat. Neurosci.* **9,** 420–428 (2006).
37. Frémaux, N. & Gerstner, W. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Front. Neural Circuits* **9,** 85 (2016).
38. Pawlak, V., Wickens, J. R., Kirkwood, A. & Kerr, J. N. Timing is not everything: neuromodulation opens the STDP gate. *Front. Synaptic Neurosci.* **2,** 146 (2010).
39. Hinton, G. The ups and downs of Hebb synapses. *PsycARTICLES* **44,** 10–13 (2003).
40. Johansen, J. P., Tarpley, J. W., LeDoux, J. E. & Blair, H. T. Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. *Nat. Neurosci.* **13,** 979–986 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements G. Venkatraman, B. Ahanonu, J. Li, B. Rossi, C. Herry, S. Ciocchi and J. Bacelo provided technical assistance. We appreciate Swiss National Science Foundation (B.F.G.), Swiss National Science Foundation, Ambizione (J.G.), US National Science Foundation (L.J.K.), Stanford University (L.J.K., J.D.M.), Simons Foundation (L.J.K.), and Helen Hay Whitney Foundation (M.C.L.) fellowships. A.L. received support from the Swiss National Science Foundation, Novartis Research Foundation, and an ERC Advanced grant. M.J.S. received support from HHMI and DARPA.

Author Contributions B.F.G. designed experiments. B.F.G., J.G.P. and J.G. established the Ca^{2+} imaging protocol and performed experiments. B.F.G., P.E.J, A.L. and M.J.S. designed analyses. B.F.G. and J.D.M. analysed data. J.A.L., L.J.K., J.D.M. and M.C.L. provided software code and advised on analyses. J.Z.L. constructed virus. F.G. and A.L. provided electrophysiological data. B.F.G. and M.J.S. wrote the paper. J.G., A.L. and all authors edited the paper. A.L. and M.J.S. supervised the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.J.S. (mschnitz@stanford.edu).

Reviewer Information *Nature* thanks V. Bolshakov and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. We housed male C57BL6/J mice (Jackson Labs; 9–10 weeks old) under a normal 12 h light/dark cycle, and provided food and water ad libitum. Before fear-conditioning experiments, we individually housed mice for at least 14 days. To habituate the mice to human handling, we handled them at least 7 times in 10 subsequent days. All animal procedures were approved and executed in accordance with institutional guidelines (Stanford Administrative Panel on Laboratory Animal Care). Mice were randomly assigned to different experimental groups in an informal manner, without regard to any of their characteristics.

Viral injection. We performed surgeries when mice were 9–10 weeks of age. We labelled excitatory neurons by injecting an adeno-associated virus (AAV, serotype 2/5) driving expression of GCaMP6m[8] via the *Camk2a* promoter. In brief, we anaesthetized mice with isoflurane (induction, 2%; maintenance, 1–2%) in 95% O_2 (Praxair) and fixed them in a stereotactic frame (Kopf Instruments). We stabilized the body temperature at 37 °C using a temperature controller and a heating pad. We injected 500 nl of the AAV (injection coordinates relative to bregma: 1.7 mm posterior; 3.4 mm lateral; 4.7 mm ventral) via a borosilicate glass pipette with a 50-μm-diameter tip using short pressure pulses applied with a picospritzer (Parker).

Microendoscope implantation. 7–12 days after AAV injection, we performed a second surgery to implant either a small custom-designed 0.6-mm-diameter microendoscope probe (Grintech GmBH) or a stainless steel guide tube (1.2 mm diameter) with a custom glass coverslip glued to one end (0.125 mm thick BK7 glass, Electron Microscopy Science). To avoid damage of the internal capsule, we chose implantation coordinates for the tip of the microendoscope that were lateral to this structure (1.7 mm posterior; lateral 3.4 mm; 4.5 mm ventral, all relative to bregma). To perform the implantation, we first made a round craniotomy centred on the injection coordinates using a trephine drill (1.0–1.8 mm in diameter). To prevent increased intracranial pressure due to the insertion of the implant, we made a circular incision in the brain to a ventral depth of 4.5 mm by using a custom-made trephine (1 mm diameter). We aspirated all brain tissue inside the trephine. Next, we lowered either the microendoscope or a metal guide tube to the bottom of the incision. We fixed the implanted guide or microendoscope to the skull using ultraviolet-light curable glue (Loctite 4305). To ensure a stable attachment of the implant, once the cranium had dried we inserted two small screws into it above the contralateral cerebellum and contralateral sensory cortex (18-8 S/S, Component Supply). We then applied Metabond (Parkell) around both screws, the implant and the surrounding cranium. Lastly, we applied dental acrylic cement (Coltene, Whaledent) on top of the Metabond, for the joint purpose of attaching a metal head bar to the cranium and to further stabilize the implant. Mice recovered for 5–7 weeks, at which point we checked the level of GCaMP6m expression using a two-photon microscope and a 20× objective lens (LUCPlan FLN, 0.5 NA, Olympus). If expression was sufficiently bright, we considered the mouse ready for mounting of the miniature microscope (nVista HD, Inscopix Inc.).

Mouse behaviour. For studies comparing a range of unconditioned stimuli (Extended Data Fig. 2), on the first day of testing we played eight sets of 10 kHz tones (85 dB, 25 tone pulses per set, each pulse 200 ms in duration, delivered at 1 Hz) while the mice were freely moving in an unfamiliar chamber. After 1 day of water restriction, we transferred mice to an experimental chamber where they received 30 μl of 4% sucrose water (500 ms reward delivery time). In the same session, after delivery of sucrose water we transferred mice to the conditioning chamber, where we delivered eight electric shocks above one eyelid (3 mA; 2 s duration) or to the paws (0.6 mA; 2 s duration) in a pseudo-random order.

Fear-conditioning experiments involved a separate cohort of mice than that used for US comparisons, and took place in two different isolation chambers, chamber A (days 1, 2, 4, 5 and 6) and chamber B (day 3). The two chambers differed in their odours, shapes, lighting pattern, and textures of the walls and floor. Before each imaging session, we cleaned the chambers with a solution of 1% acetic acid (chamber A) or 70% ethanol (chamber B). For scoring of freezing behaviour, we used video-based freezing detection software (Freeze Frame, Actimetrics) that provided a binary time trace of the mouse's movement amplitude. The video frame rate was 20 Hz, but for behavioural analysis we down-sampled the resulting time trace to 5 Hz. Mice were scored as freezing if movement was below a minimum threshold for at least 1 s. To validate the semi-automated detection of freezing, we compared freezing values to a classical time-sampling procedure during which a human observer who was blinded to the experimental conditions visually scored freezing behaviour. Freezing values with both procedures were nearly identical ($92 \pm 3\%$, $n = 12$ mice).

Throughout habituation, training and extinction sessions, the CS^+ and CS^- comprised 25, 200-ms-long tone pulses (4 kHz at 85 dB, or 10 kHz at 80 dB, with the 25 pulses delivered at 1 Hz). The acoustic frequencies of 4 kHz and 10 kHz were randomly assigned as the CS^+ and CS^- for the different mice, in a counterbalanced manner. During habituation (days 1, 2) and conditioning (day 3), mice received five CS^+ and five CS^- presentations in a pseudorandom order. During fear testing and extinction sessions (days 4–6), mice received two CS^- presentations before and two CS^- after a block of 12 unreinforced CS^+ presentations[1,41]. On all days, the inter-stimulus intervals between CS presentations were pseudo-randomly chosen between 20 and 180 s.

During conditioning on day 3, at 800 ms after the termination of the last tone pulse in each CS^+, the mouse received a US foot shock. To achieve reliable and robust fear learning, we used a relatively long (2 s) and strong foot shock (0.6 mA), which led to conditioned, CS^+-evoked freezing levels (70–90%) comparable to those reported previously for similar US parameters in mice[41]. This is a form of auditory, associative fear conditioning that is amygdala-dependent[1,12,14] (Extended Data Fig. 3) and hippocampal-independent[4]. We analysed the behavioural performance of all mice tested and retained the data regardless of freezing levels.

Behaviour controls. We examined whether microendoscope implantation in the BLA affected motor behaviour by monitoring mouse locomotion during the first two sessions in chamber A, for mice that had no, unilateral, or bilateral microendoscope implants. We used a standard video camera (AVT, GuppyPro, F125B) and the image acquisition toolbox in MATLAB to acquire movies of the mouse's behaviour at a 20 Hz frame rate. To extract the mouse's locomotor trajectory we used a custom video-tracking routine written as a plugin for the ImageJ (NIH) image analysis software environment. From these trajectories we calculated the total distance travelled, mean speed and mean acceleration (Extended Data Fig. 3a, b).

We also investigated whether microendoscope implantation affected fear learning by comparing conditioned freezing behaviours for the different groups of mice (Extended Data Fig. 3c–f). In addition to the three groups of mice used for locomotor studies, we also included a group of mice that had bilaterally implanted guide tubes through which we administered the $GABA_A$ agonist muscimol 10–15 min before the day 3 conditioning session. These metal guide tubes had the same outer diameter as the implant used for Ca^{2+} imaging, and we connected them to a 10-μl micro-syringe (Hamilton) via polyethylene (PE 20) tubing. We dissolved muscimol (Sigma-Aldrich) in artificial cerebrospinal fluid (pH 7.4) and infused this solution bilaterally into each BLA through 33-gauge infusion cannulae, each of which extended 0.5 mm beyond their corresponding metal guide tube. 10–15 min before the day 3 conditioning session, into each BLA we infused a small volume of 0.3 μl that we delivered using a syringe pump (UMP3, World Precision Instruments) at a rate of 0.2 μl min^{-1}. The infusion cannulae remained in place for 1 min after the infusion. The final dosage and volume of muscimol delivered was 2.6 nmol and 0.3 μl per BLA, as in prior fear-conditioning studies in mice[42].

Ca^{2+} imaging using the miniature microscope. We first characterized the optical working distance between the glass surface of the microendoscope inside the brain and the cells at the focal plane, by using a combination of empirical measurements and computational ray tracing simulations of the optical pathway. First, we empirically determined the distance between the back focal plane, where the image of the cells was projected outside the microendoscope, and the microendoscope's external surface protruding from the cranium. To do this, starting with the miniature microscope in a position such that the cells of interest were in focus, we lowered the microscope towards the microendoscope until we focused upon the microendoscope's external surface. After noting the distance change between these two focal positions, we used the position of the back focal plane in combination with the microendoscope's optical design to determine computationally the optical working distance between the cells and the surface of the microendoscope inside the brain. For these computations we used optical ray tracing software (Zemax). This yielded values for the working distance within the range 77–181 μm. Histological reconstructions showed that the tip of the microendoscope generally lay in the lateral amygdala. However, because the optical focal plane often spanned ventral parts of lateral amygdala and dorsal parts of the basal amygdala, we use the joint term basal and lateral amygdala (BLA) throughout.

To mount the base plate of the miniature microscope on the cranium, we attached the microscope to the base plate and lowered the pair towards the implanted microendoscope until we observed green fluorescent cells. We selected a 600 μm × 600 μm field-of-view medial from the non-fluorescent axonal fibre tract that separated the BLA and the cortex (Extended Data Fig. 1c). We glued the base plate onto the skull using blue-light curable glue (Flow-it, Pentron). Afterward, we detached the microscope and returned the mouse to its home cage.

Before each Ca^{2+} imaging session, we briefly head-fixed the mouse using its metal head-bar while allowing it to walk or run in place on a wheel. We then attached the miniature microscope to its base plate and returned the mouse to its home cage for 50–60 min until the start of the imaging session. Each session involved 22–28 min of Ca^{2+} imaging across a field-of-view of approximately 600 μm × 600 μm, which we always verified matched that seen in any previous

sessions in the same animal. After each session we detached the microscope and returned the mouse directly to its home cage for ~22 h.

To acquire fluorescence Ca^{2+} imaging videos, we used 100–150 μW of illumination intensity at the specimen and took 12 bit images (1,000 × 1,000 pixels) at a frame rate of 20 Hz. Each pixel covered 0.6 × 0.6 μm in tissue. We streamed the video data directly to hard disk (90–100 MB s^{-1}).

Two-photon imaging. To check the expression of GCaMP6m in the BLA, we used two-photon imaging to image the BLA in isoflurane-anaesthetized mice (1–2% isoflurane in O_2). We head-fixed the mice via the implanted metal head bar and positioned the implanted microendoscope under a 20× microscope objective (Olympus, LCPLFLN20xLCD) of an upright two-photon fluorescence microscope (Bruker). We first used wide-field epi-fluorescence imaging to visualize the BLA tissue through the microendoscope. We then switched to two-photon laser-scanning imaging and acquired images of 256 × 256 pixels at a 3 Hz frame rate.

Basic processing of the Ca^{2+} imaging videos. To account for slowly varying illumination non-uniformities across the field-of-view, we normalized each image frame by dividing it by a spatially low-pass filtered (length constant: 120 μm) version of the frame using ImageJ software (NIH). Next, we used the ImageJ plugin TurboReg[43] to correct for lateral motions of the brain by performing a rigid image registration across all frames of the movie. After motion correction, we temporally smoothed and down-sampled each movie from 20 Hz to 5 Hz. We then re-expressed each image frame in units of relative changes in fluorescence, $\Delta F(t)/F_0 = (F(t) - F_0)/F_0$, where F_0 is the mean image obtained by averaging the entire movie.

Cell sorting. We identified spatial filters corresponding to individual neurons using an established, automated cell sorting routine based on principal and independent component analyses[7,44]. As in previous Ca^{2+} imaging studies using the miniature microscope[7,45], after motion correction we identified cells' spatial filters based on the Ca^{2+} data acquired over the entire session. For each filter, we then zeroed all pixels with values <50% of that filter's maximum intensity. To obtain time traces of Ca^{2+} activity, for each cell we applied its thresholded spatial filter to the $\Delta F(t)/F_0$ movie.

As previously described[44], the extracted spatial filters generally had sizes, morphologies and activity traces that were characteristic of individual neurons, but there were also some spatial filters that were obviously not neurons and that we discarded (Extended Data Fig. 2b). For the 4–10% of candidates with less common characteristics, we were conservative and accepted only those that were plainly cells by human visual scrutiny. We verified every cell included in the analyses by visual inspection.

Registration of cell identities across imaging sessions. We generated cell maps for each day by projecting thresholded versions of the spatial filter of each cell onto a single image[7] (Extended Data Fig. 5a). Taking the map from day 3 as a reference, we aligned the other cell maps to this one via a scaled image alignment using the TurboReg image registration algorithm[43]. This corrected slight translations, rotations, or focus-dependent magnification changes between sessions and yielded the location of each cell in the reference coordinate system.

We then identified candidate cells across sessions that might be the same neuron seen on multiple occasions. To do this, we applied the observations that our image registration procedure had sub-micrometre precision, and that the distance between the centroids of neighbouring somata was always >6 μm (Extended Data Fig. 5d). We thus enforced that all observed cells deemed to be the same neuron had all pair-wise separations ≤6 μm (Extended Data Fig. 5e). The distribution of pair-wise separations between cells assigned the same identity yielded the conservative estimate that 99.7% of these assignments were correct (Extended Data Fig. 5e, inset).

Identification of neuronal sub-classes. We identified functional sub-classes of neurons by comparing the stimulus-evoked fluorescence Ca^{2+} signals of individual cells to their baseline fluorescence levels, using 1 s time bins. To compute the baseline activity level of each cell, we averaged its fluorescence signal over the complete imaging session excluding all stimulus presentations. For the analyses of neural responses to a CS^- or CS^+ (always in the form of 25 tone pulses, 200 ms in duration, delivered at 1 Hz), we defined the stimulus response period as the 25-s-period that began at the onset of the first tone pulse and extended 800 ms beyond the offset of the twenty-fifth pulse (that is, up to the start of the US). To analyse neural responses to a shock US, we defined the stimulus response period as the 2 s period of eyelid or foot shock delivery. To analyse the neural responses to sucrose water, we defined the stimulus response period as the 1 s interval starting from the onset of stimulus delivery. After computing stimulus-evoked fluorescence responses of each cell in 1-s time bins, we compared the set of all such responses to the baseline activity level of each cell using the Wilcoxon rank-sum test. All cells with stimulus-evoked responses that were significantly different from baseline activity (significance criterion: $P \leq 0.01$) were classified as CS- or US-responsive.

We also verified that the definition of baseline activity had little effect on the sets of cells identified as having stimulus-evoked responses, by comparing the results obtained using two different definitions. In one case, we determined the level of baseline activity of each cell by finding its average activity across the full imaging session, excluding stimulus presentations. Alternatively, we used the 20-s-period immediately before each stimulus presentation to assess the magnitude of the stimulus-evoked response. Using all 3,655 cells for this validation analysis, we found that 3,524 cells (96%) were categorized identically under the two definitions of baseline activity, indicating that the choice of definition had little effect on our subsequent analysis results.

To identify neurons that significantly increased or decreased their CS-evoked activity during the five paired CS–US presentations on day 3, we compared their CS-evoked Ca^{2+} signals for CS presentations early in the session (presentations 1 and 2) versus those late in the session (presentations 3–5) (Wilcoxon rank-sum test, using a significance threshold of $P \leq 0.05$). To identify cells with significantly increased or decreased CS-evoked activity after conditioning, we compared CS-evoked Ca^{2+} signals from the days before (days 1, 2) and after (days 4–6) the training session on Day 3 (Wilcoxon rank-sum test, using a significance threshold of $P \leq 0.05$).

Population vector analyses. We analysed our data with MATLAB (Mathworks) using the image processing and machine learning toolboxes. For population vector analysis, decoder training and testing we used neuronal Ca^{2+} signals expressed as relative fluorescence changes ($\Delta F/F$), down-sampled the traces to 1 s time bins, and organized the data to contain equal numbers of time points for baseline, CS^+, CS^- or US presentations. We chose 1 s bins, because this choice yielded superior decoding performance compared to the use of either shorter or longer time bins. To quantify the similarity of two sets of neuronal ensemble response patterns, we calculated the Mahalanobis distances between the two sets of population activity vectors[21]. To do this, we created a group of n-dimensional (n = number of neurons) activity vectors, x, for each stimulus type (baseline, CS^-, CS^+ or US) and calculated the PVD between the two groups (Extended Data Fig. 10). For example, the Mahalanobis PVD (M) between sets of CS- and US-evoked ensemble activity patterns is:

$$M^2 = (x - \mu)^{\mathrm{T}} \cdot \Sigma^{-1} \cdot (x - \mu)$$

where x and μ are individual and mean population vectors for CS and US ensemble responses, respectively, and x^{T} and μ^{T} are their transposes. $\sum$ is the covariance matrix for the set of ensemble responses. The Mahalanobis distance takes into account the differences in the means of the two sets of ensemble responses as well as their co-variances (Extended Data Fig. 9). We averaged the PVD over all points x in both sets of ensemble responses.

To track the CS–US PVDs across the day 3 training session, we down-sampled all neural activity traces to 1-s time bins. This resulted in 25 time bins for each 25-s CS presentation and 2 time bins for each 2-s US presentation. Next, we constructed the mean CS^+, CS^- and US population vectors by averaging the evoked neural responses over all five presentations of each stimulus and all the time bins associated with each stimulus presentation. To calculate the change in the CS^+ population vector expected under a cellular, Hebbian model of associative potentiation, we restricted the changes to the CS^+ population vector to those cells that were US-responsive and used the empirically determined mean stimulus-evoked responses of these cells to calculate the vector entries.

Decoding ensemble neural activity. We constructed all binary (Fig. 3b; Extended Data Fig. 8) and three-way (Fig. 3a; Extended Data Fig. 7) Fisher linear decoders[21] in MATLAB. To construct the three-way decoders, we used an established approach based on multiclass Fisher linear discriminant analysis that maximizes the ratio of the mean variances between the different classes to that within the individual classes[21]. We used the set of neural ensemble Ca^{2+} response traces ($\Delta F/F$) from each mouse and trained decoders to discriminate the Ca^{2+} activity patterns that occurred during baseline epochs, CS^+ or CS^- presentations. Before training we down-sampled the data into 1-s time bins. We determined decoder performance values as the mean rate of correct predictions over a tenfold cross-validation. For cross-validation, we split each dataset into 10 equally sized blocks and randomly assigned each time bin to one of the 10 blocks; we used 9 of the blocks for decoder training and 1 for testing. To evaluate the statistical significance of decoding performance, we trained control decoders on temporally shuffled datasets, and compared the mean, cross-validated performance values to those of the real decoders.

When making comparisons across decoders involving unequal numbers of cells (Fig. 3a), we confirmed all results via a control analysis that used statistical re-sampling methods[46] to construct decoders based on equal numbers of cells; this yielded decoding results virtually indistinguishable from those shown in Fig. 3a. As a further check, we also verified that the small performance difference between decoders based on all cells and those based only on CS-responsive neurons was not

simply due to the smaller number of cells used for the latter decoders, as opposed to the information content of their activity traces. For this purpose, we constructed control decoders (Fig. 3a, dashed green curve) based on the same number of cells as used for the decoders of CS-responsive cells, but with the cells randomly chosen. The accuracy difference between these control decoders and that of the decoders of CS-responsive neurons was notable, as the control decoders performed at levels very close to chance and no better than decoders based on temporally shuffled neural activity traces (Fig. 3a, dashed grey curve).

Construction of the CS$^+$ rescue decoder. We constructed and validated the rescued time–lapse CS$^+$ decoder in five main steps (Extended Data Fig. 10a, b). Step 1: we recorded CS$^+$ ensemble activity before conditioning (days 1 and 2). Step 2: we recorded neuronal population activity during conditioning with five CS–US paired presentations (day 3) and identified individual neurons that altered their CS$^+$-evoked responses (Wilcoxon rank-sum test, comparing CS$^+$-evoked responses between the early (CS–US pairings 1 and 2) and late phase (CS–US pairings 3–5) of conditioning, significance threshold $P < 0.15$). Step 3: we simulated the full, consolidated CS$^+$ ensemble transformation by gradually extrapolating changes of individual neuron responses and adding them to their CS$^+$ responses before conditioning (Extended Data Fig. 10c). Step 4: we trained a new rescue decoder and evaluated its performance for different extrapolation magnitudes. Step 5: to validate the simulated transformation of ensemble coding, we compared the average performance of the rescue decoder to the average performance of the stable CS$^-$ time-lapse decoder.

Relating neural population vectors to freezing behaviour. To examine how ensemble neural activity related to each mouse's overall level of conditioned freezing (Fig. 5a), we first calculated for each individual CS$^+$ (or CS$^-$) presentation the PVD to the mean US population vector, and then normalized the resulting CS–US PVD by the value of the CS–US PVD computed for the mouse's first CS$^+$ (or CS$^-$) presentation. We plotted these normalized CS–US PVD values as a function of the percentage of time during each 25-s CS presentation that the mouse spent freezing (Fig. 5a).

To examine whether BLA ensemble neural activity differed between the moments within individual CS$^+$ presentations when a mouse was freezing versus not freezing (Fig. 5b), we divided each 25-s CS$^+$ presentation into 1-s time bins. For each CS$^+$ presentation we then found the ratio of the CS$^+$–US PVDs, as computed for the 1-s time bins when the mouse was freezing versus those when the mouse was not freezing. We plotted this ratio as a function of the proportion of time during the 25-s CS$^+$ that the mouse spent freezing (Fig. 5b).

Next, we examined how the change in each mouse's CS$^+$–US PVD during learning related to the change in its freezing behaviour (Fig. 5c). For each mouse we computed the percentage change in the CS$^+$–US PVD occurring between the last six CS$^+$ presentations before learning (days 1 and 2) and the first six CS$^+$ presentations after learning (day 4). We plotted the resulting values versus the changes in freezing behaviour across the same time periods.

We performed a similar analysis to examine how the change in each mouse's CS$^+$–US PVD during extinction training related to the consolidated change in its freezing behaviour (Fig. 5d). We compared the first six CS$^+$ presentations from the first day of extinction training (day 4) to the first six CS$^+$ presentations on the last day of extinction learning (Day 6). For each mouse we computed the percentage differences in CS$^+$–US PVDs across these two groups of CS$^+$ presentations, and we compared the resulting values to the ratio of the mouse's freezing levels during these two sets of CS$^+$ presentations.

Histological verification of cell identity. Four weeks after injection of the GCaMP6m viral construct or two weeks after the imaging experiments, we transcardially perfused mice with PBS followed by ice-cold 4% paraformaldehyde (PFA). Next, we extracted mouse brains and kept them for post-fixation in PFA for 24–48 h. We then cut 100-μm-thick coronal brain slices using a Vibratome (VT1000 s, Leica) and stored all slices in PBS.

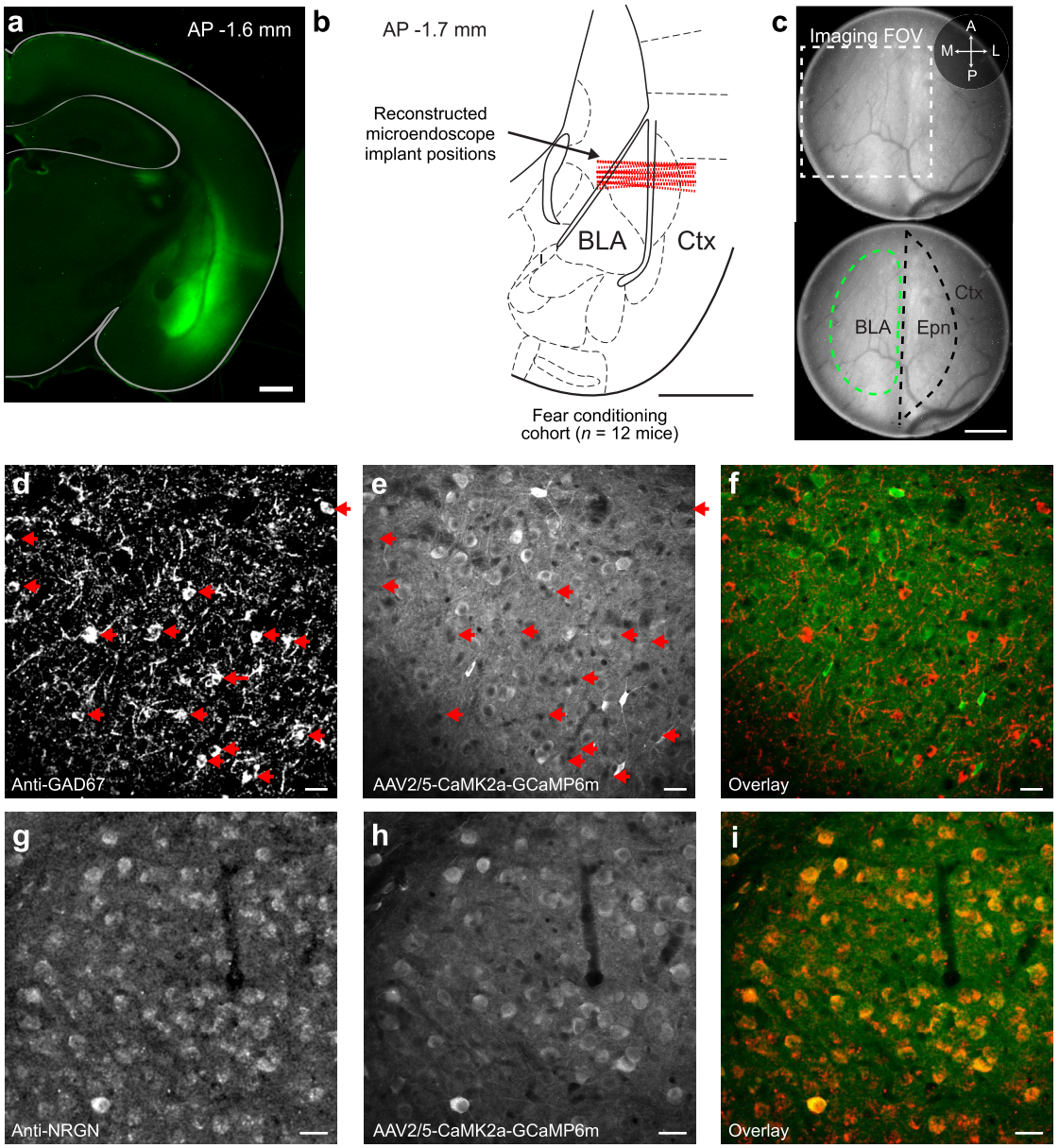
To validate the implant positions of the microendoscopes relative to the BLA we mounted all coronal brain slices on microscopy slides and acquired large field-of-view fluorescence images using a standard fluorescence macroscope (Z16, Leica). We then overlaid all images with a validated reference image[47], marked the position of the endoscope tip for every mouse (Extended Data Fig. 1b), and determined the ventral depth of the implant with respect to bregma, using the coordinate system of the reference image.

To stain inhibitory or excitatory neurons, we followed standard immunostaining procedures. In brief, we incubated brain slices with the primary antibodies, rabbit anti-GAD65 (1:500, AB1511, EMD Millipore) or anti-Neurogranin (1:10,000, 07-425, EMD Millipore) at 4 °C overnight followed by a second overnight incubation at 4 °C with secondary anti-rabbit Alexa 647 antibodies (1:1,000, both Invitrogen).

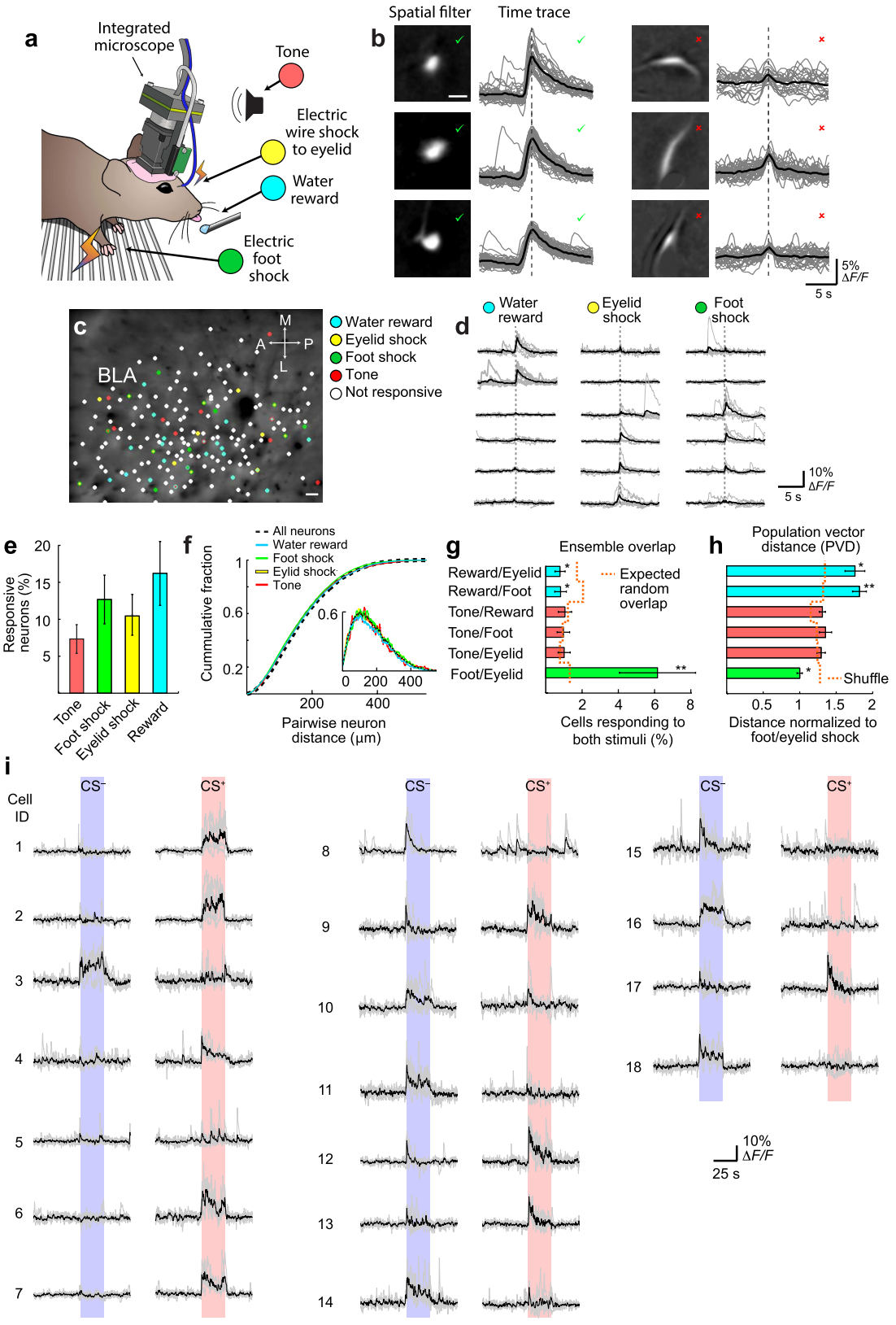
Data analyses and statistical tests. We conducted all analyses using custom routines written in MATLAB (Mathworks) and ImageJ (NIH) software. Two-tailed, non-parametric statistical tests were used throughout to avoid assumptions of normal distributions and equal variance across groups. All signed-rank tests were Wilcoxon signed-rank tests. All rank-sum tests were Wilcoxon rank-sum tests. For ANOVA, the Friedman and Kruskal–Wallis tests, respectively, were used for one-way ANOVAs with and without repeated measures. Supplementary Table 1 summarizes the results from these ANOVA analyses. The sizes of our mice samples were chosen to approximately match those of previous work, as there was no pre-specified effect size. Investigators were not blinded to an animal's experimental cohort.

Code and data availability. The algorithm used for image registration is available on its author's website[43]. The algorithm used for cell sorting is available as published supplementary material[44]. Other software code and the data that support the findings of this study are available from the corresponding author upon reasonable request.

41. Courtin, J. *et al.* Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature* **505,** 92–96 (2014).
42. Raybuck, J. D. & Lattal, K. M. Double dissociation of amygdala and hippocampal contributions to trace and delay fear conditioning. *PLoS One* **6,** e15982 (2011).
43. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7,** 27–41 (1998).
44. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63,** 747–760 (2009).
45. Betley, J. N. *et al.* Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature* **521,** 180–185 (2015).
46. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497,** 585–590 (2013).
47. George Paxinos, K. B. J. F. *The Mouse Brain in Stereotaxic Coordinates* Vol. 2 (2001).
48. Singec, I., Knoth, R., Ditter, M., Volk, B. & Frotscher, M. Neurogranin is expressed by principal cells but not interneurons in the rodent and monkey neocortex and hippocampus. *J. Comp. Neurol.* **479,** 30–42 (2004).
49. Maren, S., Yap, S. A. & Goosens, K. A. The amygdala is essential for the development of neuronal plasticity in the medial geniculate nucleus during auditory fear conditioning in rats. *J. Neurosci.* **21,** RC135 (2001).

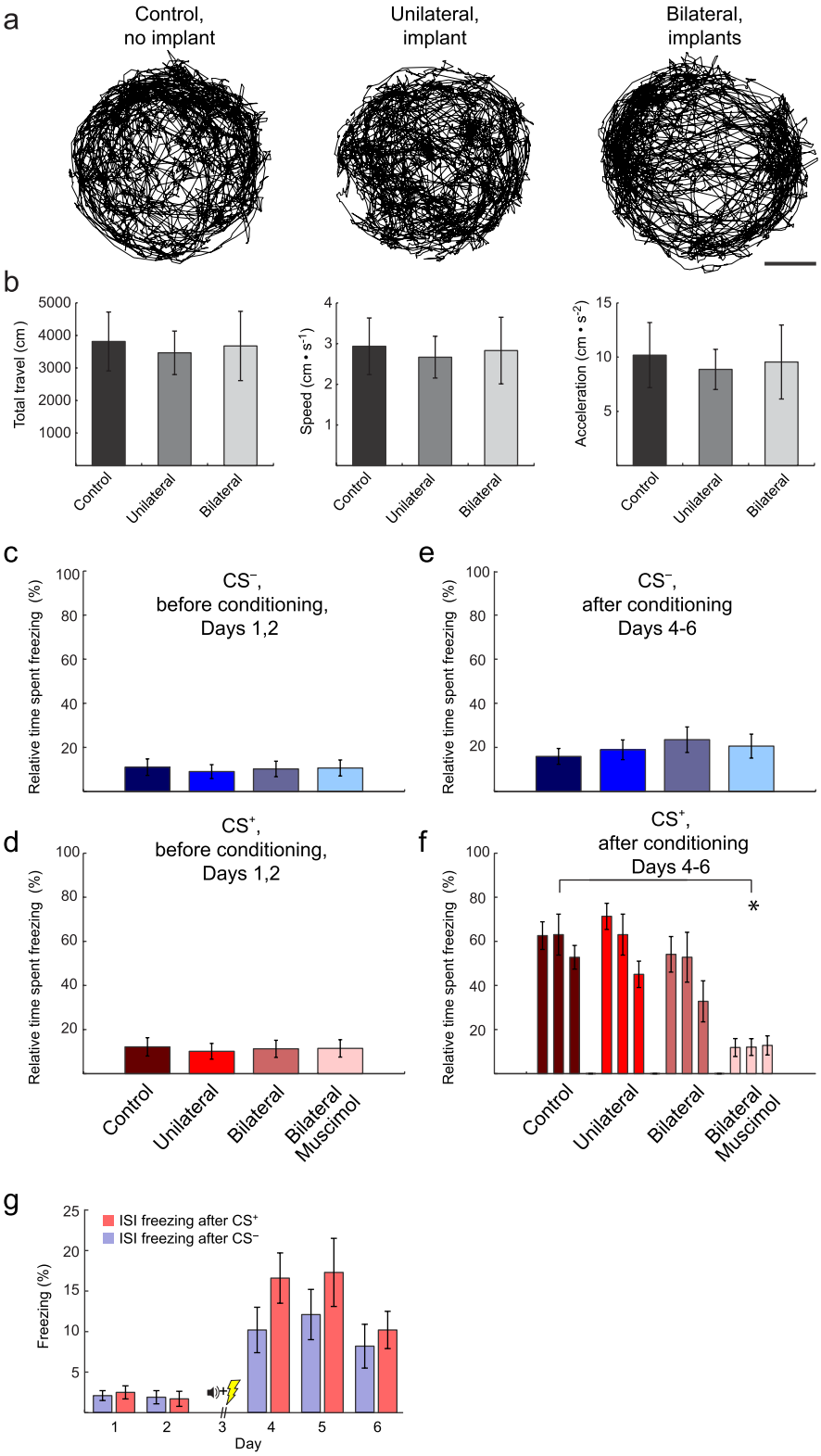


Extended Data Figure 1 | Mouse preparation for Ca^{2+} imaging in excitatory BLA neurons. a, Coronal slice of a mouse brain showing expression in the BLA of the GCaMP6m Ca^{2+} indicator, 5 weeks after injection of the AAV2/5-CaMK2a-GCaMP6m virus. Scale bar, 1 mm. **b**, Schematic of a coronal mouse brain section shown with the reconstructed positions (dashed red lines) of the microendoscope implants in the BLA, for the 12 mice subject to the experimental protocol of Fig. 1c. The focal planes for *in vivo* Ca^{2+} imaging were 77–181 μm below the indicated implant positions, as determined by computational modelling of the microendoscope optical pathway, using the empirically determined value of the back focal length. The optical focal plane often spanned ventral parts of lateral amygdala and dorsal parts of the basal amygdala, hence the use of the term basal and lateral amygdala (BLA) throughout. AP, anterior posterior; Ctx, piriform cortex. Scale bar, 1 mm. The mouse brain section has been reproduced with permission from ref. 47. **c**, Top, wide-field fluorescence image of BLA tissue acquired through an implanted microendoscope, 6 weeks after injection of the AAV2/5-CaMK2a-GCaMP6m virus. The outer fibre tract enclosing the BLA does not express GCaMP6m and appears as a vertical dark stripe in the centre of the field-of-view. The dashed box shows the position of the camera's field-of-view, which we positioned over the BLA by using the fibre tract as a reference marker. Bottom, the same image but with the boundaries of the BLA and endopiriform nucleus (Epn) marked in green and black dashed lines, respectively. Scale bar, 0.2 mm. **d**–**f**, Coronal section of a mouse brain showing, inhibitory neurons in the BLA immuno-labelled with a monoclonal anti-GAD67 antibody (**d**), neurons expressing GCaMP6m under the control of the *Camk2a* promoter (**e**) and the overlay of the images in **d** and **e** (**f**). Red arrows in **d** and **e** mark $GAD67^+$ interneurons that are not expressing GCaMP6m. Scale bars, 20 μm. **g**–**i**, Coronal brain section showing excitatory neurons in the BLA immunolabelled using a polyclonal anti-neurogranin (NRGN) antibody[48] (**g**), neurons expressing GCaMP6m (**h**) and an overlay of the images in **g** and **h** (**i**), showing that the set of $NRGN^+$ excitatory neurons (labelled red) strongly overlap with the set of cells expressing GCaMP6m (labelled green). Scale bars, 20 μm.



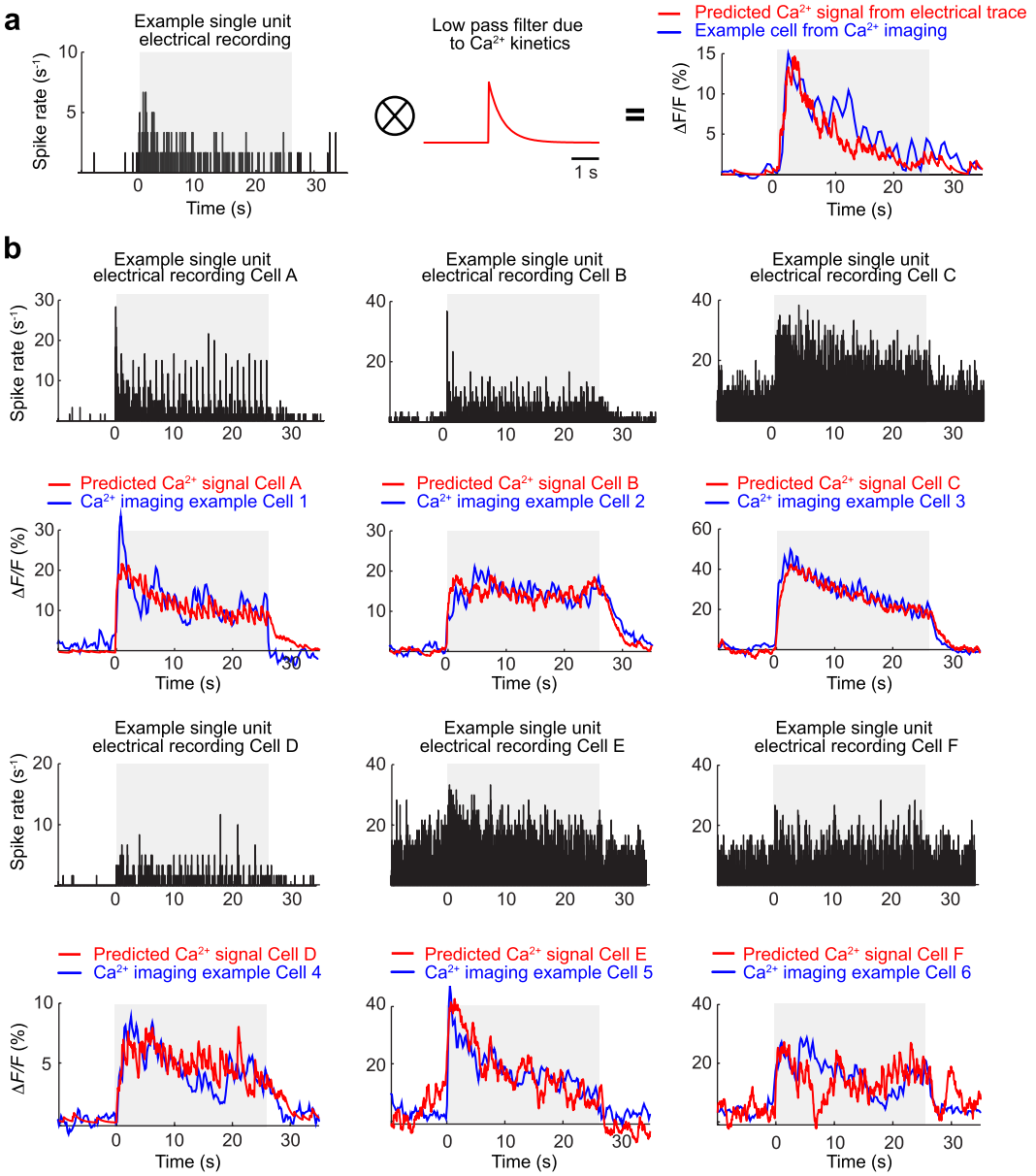
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Stimuli of neutral, positive, and negative valence activate sparse, largely distinct, spatially intermingled subsets of neurons in the BLA. a, A miniature fluorescence microscope enabled large-scale neural Ca^{2+} imaging in the BLA of awake behaving mice as we presented stimuli of different valences to the animals. **b**, Candidate cells identified using an automated cell sorting routine[7,44] were easily segregated into those with sizes, morphologies and Ca^{2+} activity traces (grey traces, individual activity transients; black traces, mean waveforms) characteristic of individual neurons (left), and those that were obviously not neurons (right). For the 4–10% of candidates with less common characteristics, we accepted only those that were plainly cells by human visual scrutiny. We verified every cell included in the analyses by visual inspection. **c**, An example cell map in the BLA, as determined from a Ca^{2+} imaging dataset 28 min in duration. Colours indicate subsets of BLA neurons that responded significantly to rewards (light blue), electric foot shocks (green), eyelid shocks (yellow), or neutral tones (red). Scale bar, 20 μm. $P \leq 0.01$, rank-sum test, comparing evoked Ca^{2+} signals to baseline levels. **d**, Ca^{2+} responses of six example neurons in the same mouse after the delivery of individual water rewards (left), eyelid shocks (middle) or foot-shocks (right). Grey traces show Ca^{2+} responses from eight individual trials. Black traces show the mean responses. **e**, Percentages of cells ($n = 1{,}251$ neurons in total from 8 mice) with significant Ca^{2+} responses to the four different stimuli (threshold for a significant response: $P \leq 0.01$, comparing evoked versus baseline Ca^{2+} levels for $n = 8$ presentations of the stimulus; Wilcoxon rank-sum test). Error bars denote s.e.m. **f**, Cumulative probability distributions, each determined as a mean over 8 mice (1,251 total cells), of the centroid separations between all pairs of cells in each mouse (dashed black curve), and between pairs of cells that both had significant responses to one of the four different stimuli (coloured curves). Inset, the corresponding probability densities. **g**, Percentages of all neurons ($n = 8$ mice; 1,251 cells in total) that had significant responses to each of the two stimuli in each pair listed on the vertical axis. Dashed orange line indicates expected levels of overlap due to random chance. $^{*}P < 0.05$, $^{**}P < 0.01$, comparing actual percentages versus those determined from datasets in which we randomly shuffled the identities of the cells (1,000 random shuffles; Wilcoxon signed-rank test). **h**, Mahalanobis PVDs between the ensemble neural representations of the two stimuli of each pair listed on the vertical axis. All PVDs are normalized to the PVD between the neural representations of eyelid-shock and foot-shock. Pairs of stimuli with smaller PVDs have ensemble neural representations of greater similarity than pairs with larger PVDs. Dashed orange line indicates PVDs between ensembles in which we randomly shuffled the identities of the cells (1,000 random shuffles). $^{*}P < 0.05$, $^{**}P < 0.01$, comparing actual PVD values to those determined for the shuffled datasets, Wilcoxon signed-rank test. Data are based on the same 1,251 cells as in **e**–**g**. Error bars in **g** and **h** denote s.e.m. **i**, 18 sets of fluorescence Ca^{2+} traces, showing evoked responses to presentations of the CS^{+} and CS^{-} from 18 example neurons before fear conditioning. Light grey traces show the individual responses of the cells to five stimulus presentations; black traces are average responses. Traces were downsampled to 5 Hz to aid visualization.

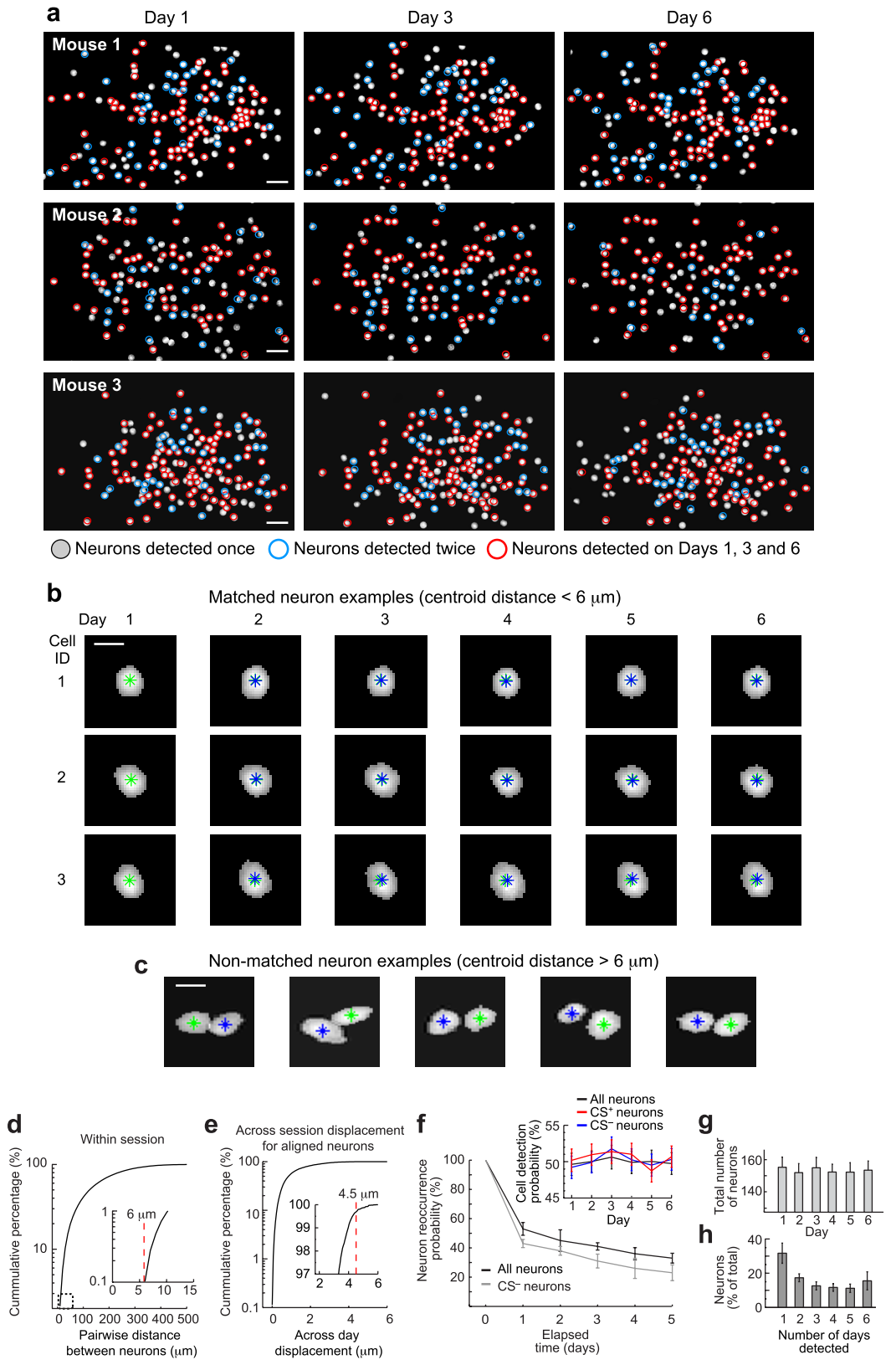


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Unilateral implantation of a microendoscope does not alter conditioned freezing; bilateral implantation minimally alters conditioned freezing without affecting locomotion. a, Traces of locomotor activity across an entire (22 min) habituation session (day 1), for one example mouse in each of the three experimental groups indicated. Scale bar, 5 cm. **b**, Total distance travelled (left), locomotor speed (middle) and acceleration (right) for the three groups of mice during the day 1 habituation session. No significant differences between the three experimental groups (no-implant control (12 mice); unilateral implant (12 mice); bilateral implant (10 mice)) (one-way Kruskal–Wallis test; degrees of freedom: $df_{group} = 2$, $df_{err} = 31$, $df_{total} = 33$ for 3 groups and 34 total mice; $\chi^2 = 10$–12; $P \geq 0.05$ for all three locomotor parameters). **c**, **d**, Time mice spent freezing before conditioning (days 1, 2) in response to 5 presentations of the CS^- (**c**) and the CS^+ (**d**), in no implant controls (12 mice), unilateral implant (12 mice), bilateral implant (10 mice), and bilateral implant plus muscimol BLA injection before the day 3 conditioning session (8 mice). No significant differences in freezing time (one-way Kruskal–Wallis test; $df_{group} = 3$, $df_{err} = 42$, $df_{total} = 45$ for 4 groups and 42 total mice; $\chi^2 = 10.2$ and 11.8 for CS^+ and CS^-, $P \geq 0.05$ for both CS^+ and CS^-). **e**, **f**, Time mice spent freezing after conditioning (days 4–6) in response to 4 presentations of the CS^- (**e**) and during 3 sets each comprising 4 presentations of the CS^+ (**f**) in the same 42 mice as in **c** and **d**. $^*P = 0.005$ (Wilcoxon signed-rank test; bilateral muscimol group versus control; significance threshold = 0.02 after Dunn–Šidák correction for 3 comparisons). Data are consistent with a study showing the necessity of BLA for auditory fear conditioning[49] and further demonstrate that the BLA we are imaging are functional and necessary for the behaviour. **g**, Time mice ($n = 12$) spent freezing during the 20–180 s inter-stimulus intervals (ISI) after either a CS^+ or CS^- presentation. CS^+ and CS^- freezing values are averages over the numbers of stimulus presentations shown in Fig. 1c. After fear conditioning, CS^--evoked freezing levels were above those during the inter-stimulus intervals, indicating the CS^- did not serve as a learned safety signal. Data in **b**–**g** are mean ± s.e.m.

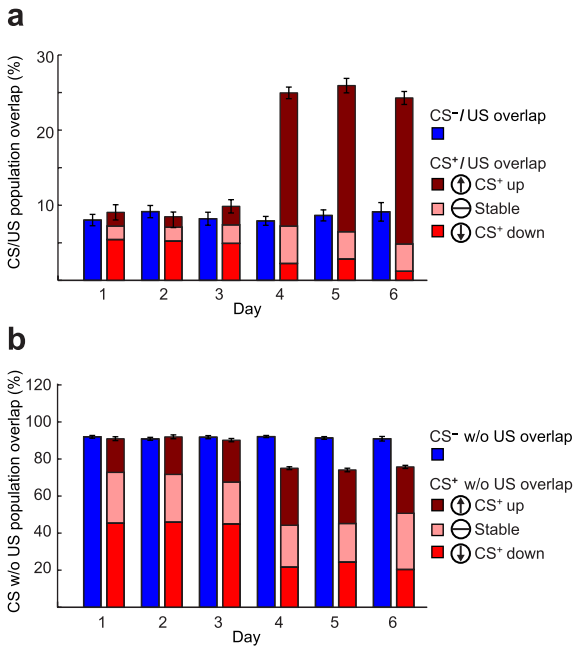


Extended Data Figure 4 | Ca^{2+} transient responses of individual BLA neurons to CS presentations closely resemble expectations based on electrical recordings of these responses. To assess whether fluorescence Ca^{2+} imaging in the BLA captured similar forms of neural activity as previous extracellular electrical recordings in this brain area, we compared the responses of individual neurons to CS presentations, as observed using the two recording modalities in two different sets of mice presented with the same CS stimuli. Across the two datasets, there was close agreement between the shapes of the empirically determined Ca^{2+} transient waveforms and the expected waveforms based on the electrically recorded CS-evoked spiking responses. **a**, We took a recording of CS-evoked spiking activity from an individual BLA cell (left), convolved the spike train with a decaying exponential function (700 ms time constant) to account for the kinetics of the GCaMP6m indicator (middle), and subtracted the baseline fluorescence level to yield a predicted CS-evoked Ca^{2+} fluorescence signal ($\Delta F/F$) whose waveform shape closely matched the actual CS-evoked Ca^{2+} fluorescence signal of a BLA cell that we had monitored using the miniature microscope (right). **b**, Six additional examples of the CS-evoked spiking responses in individual BLA neurons, as monitored via extracellular electrical recordings (black traces). From these spike trains, the same approach as shown in **a** was used to predict the Ca^{2+} fluorescence signals that these cells would produce (red traces), and these predictions were compared to the actual CS-evoked Ca^{2+} fluorescence signals of another six BLA cells studied by Ca^{2+} imaging with similar responses (blue traces).

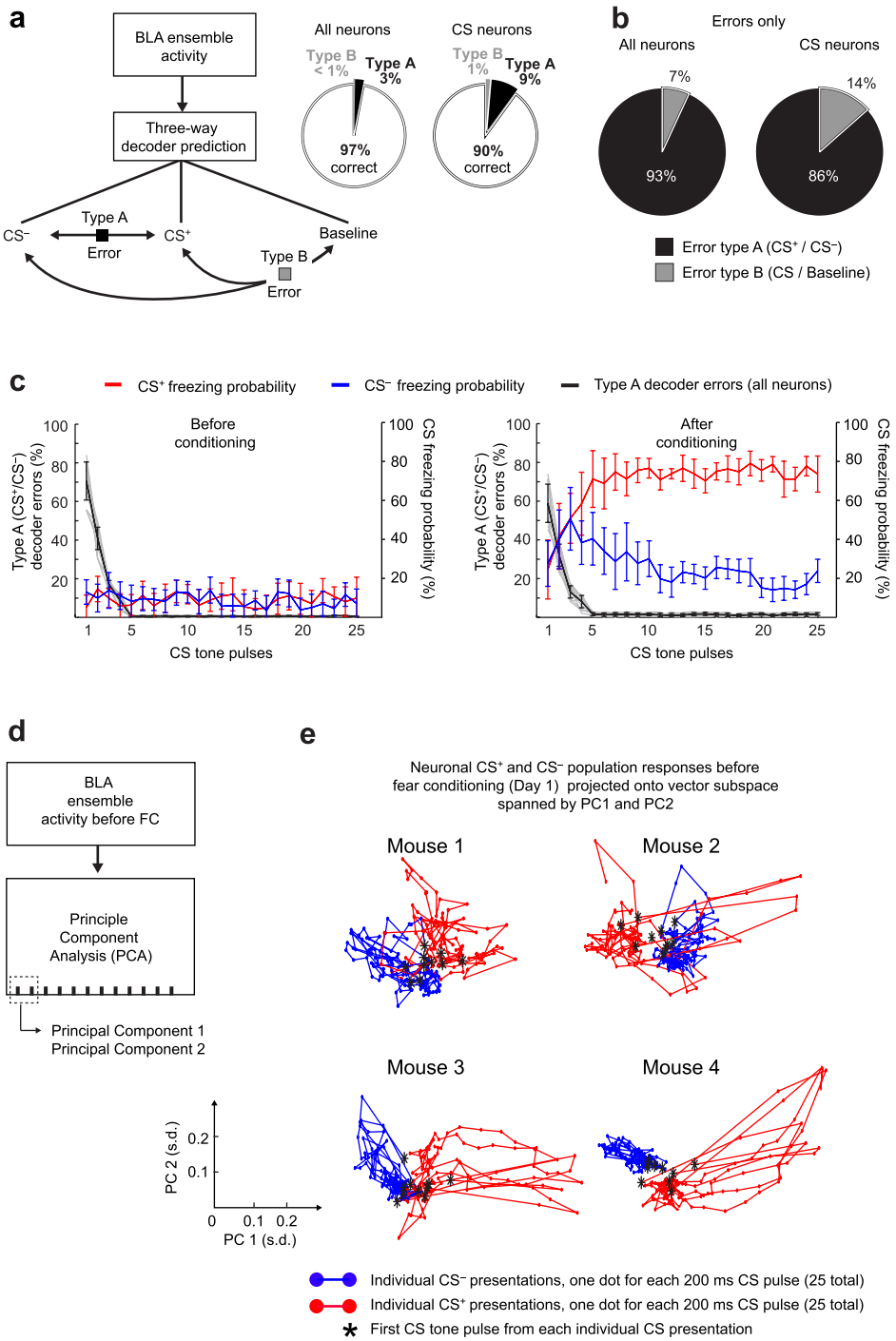


Extended Data Figure 5 | See next page for caption.

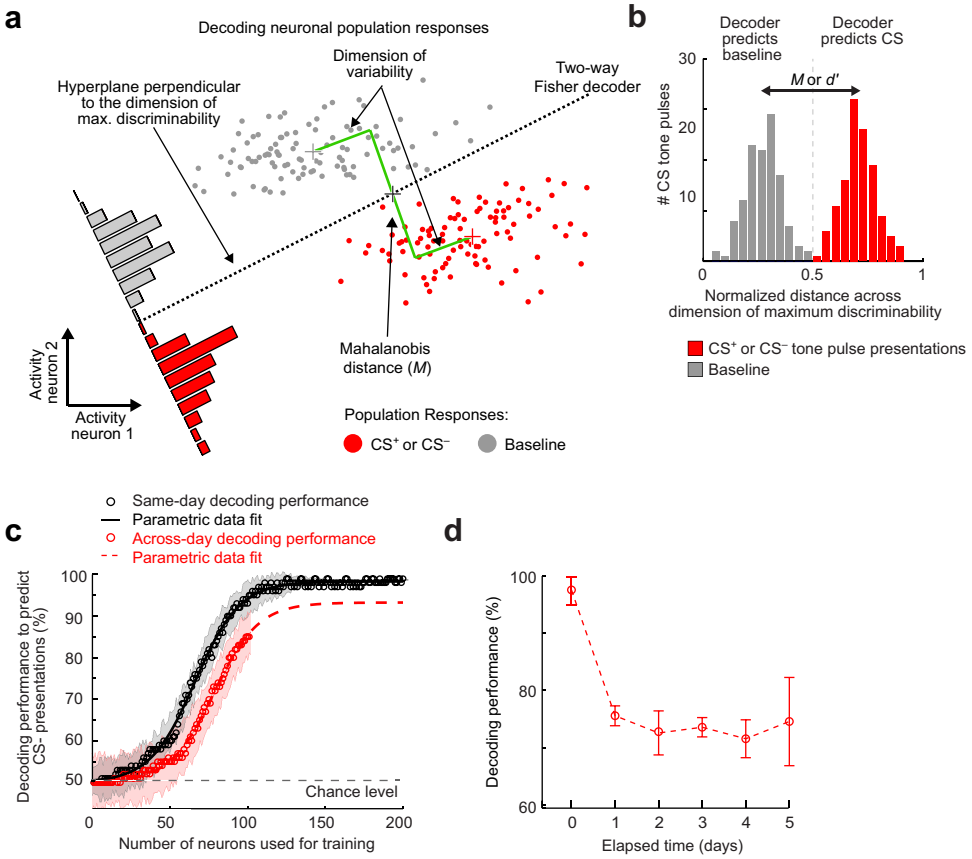
Extended Data Figure 5 | Precise spatial registration of the Ca^{2+} imaging datasets from different behavioural sessions allows unambiguous tracking of individual cells across multiple days. a, Using the spatial filters provided for each neuron by the automated cell sorting algorithm[7,44], maps of all active cells detected in the BLA on each day of the study were made. Standard methods of image alignment[43] were used to register these maps across the different days. Approximately 50% of all neurons observed across the entire experiment were detected as active on individual days. **a**, Example maps of active BLA cells from three mice on the first (left), third (middle), and last (right) day of the 6-day experimental protocol (Fig. 1c). Circles indicate cells that were active in only one of the three recordings (grey), on two of the three days (blue), or on all three days (red). Scale bars, 30 μm. The maximum number of active cells seen in one session was 192. **b**, Thresholded spatial filters from three example cells registered across the 6-day experimental protocol. Green asterisks indicate the centroid position of each cell on day 1. Blue asterisks mark the centroid positions on subsequent days. Scale bar, 10 μm. **c**, Five examples of neighbouring cells detected via their activity patterns on different days. Two individual cells are clearly discernible in each case. Scale bar, 10 μm. **d**, Cumulative histogram of distances between the centroids of all pairs of cells detected within the same imaging session, plotted with a logarithmic scale. Inset, magnified view of the dashed box area. No pairs of cells were separated by <6 μm. **e**, Cumulative histogram of distances between the centroids of all pairs of active cells registered as being the same neuron seen in different sessions. Inset, magnified view for y-axis values $>97\%$. Because the worst-case alignment error of the image registration algorithm was 1.5 μm, as determined by bootstrap analysis[7], and as all pairs of anatomically distinct cells were separated by ≥ 6 μm (**d**), cell pairs separated by <4.5 μm were nearly guaranteed to be the same neuron seen on two different occasions. This yielded the worst-case estimate that $>99.7\%$ of all cell pairs registered as being the same cells were correctly assigned the same identity. This estimate is conservative in that the image registration errors were usually <1 μm. **f**, Probability that an active neuron detected in one imaging session will also be active in a subsequent session, for all 3,655 neurons (black) and for CS^--responsive neurons (grey). Inset, probability that a cell detected on any day in the study was present in each of the imaging sessions, for all 3,655 neurons (black), the CS^+-responsive neurons (red), and the CS^--responsive neurons (blue). These probabilities were constant throughout and statistically indistinguishable between the three groups of cells examined for all days and all mice ($49 \pm 2\%$ of all cells were active each day; two-way repeated measures ANOVA; degrees of freedom: $df_{mice} = 11$, $df_{group} = 2$, $df_{interaction} = 22$, $df_{err} = 180$, $df_{total} = 215$ for 6 days, 3 groups of cells and 12 mice; $\chi^2 = 0.7$–1.4; $P > 0.05$ for all). **g**, Number of neurons detected in each mouse was stable across all days (152 ± 14 cells per day; $n = 12$ mice; one-way Friedman test; $df_{days} = 5$, $df_{err} = 55$, $df_{total} = 71$ for 6 days and 12 mice; $\chi^2 = 5.9$; $P = 0.31$). **h**, Percentages of all 3,655 cells in the study that were detected in 1–6 sessions. Data in **f**–**h** are mean ± s.e.m.



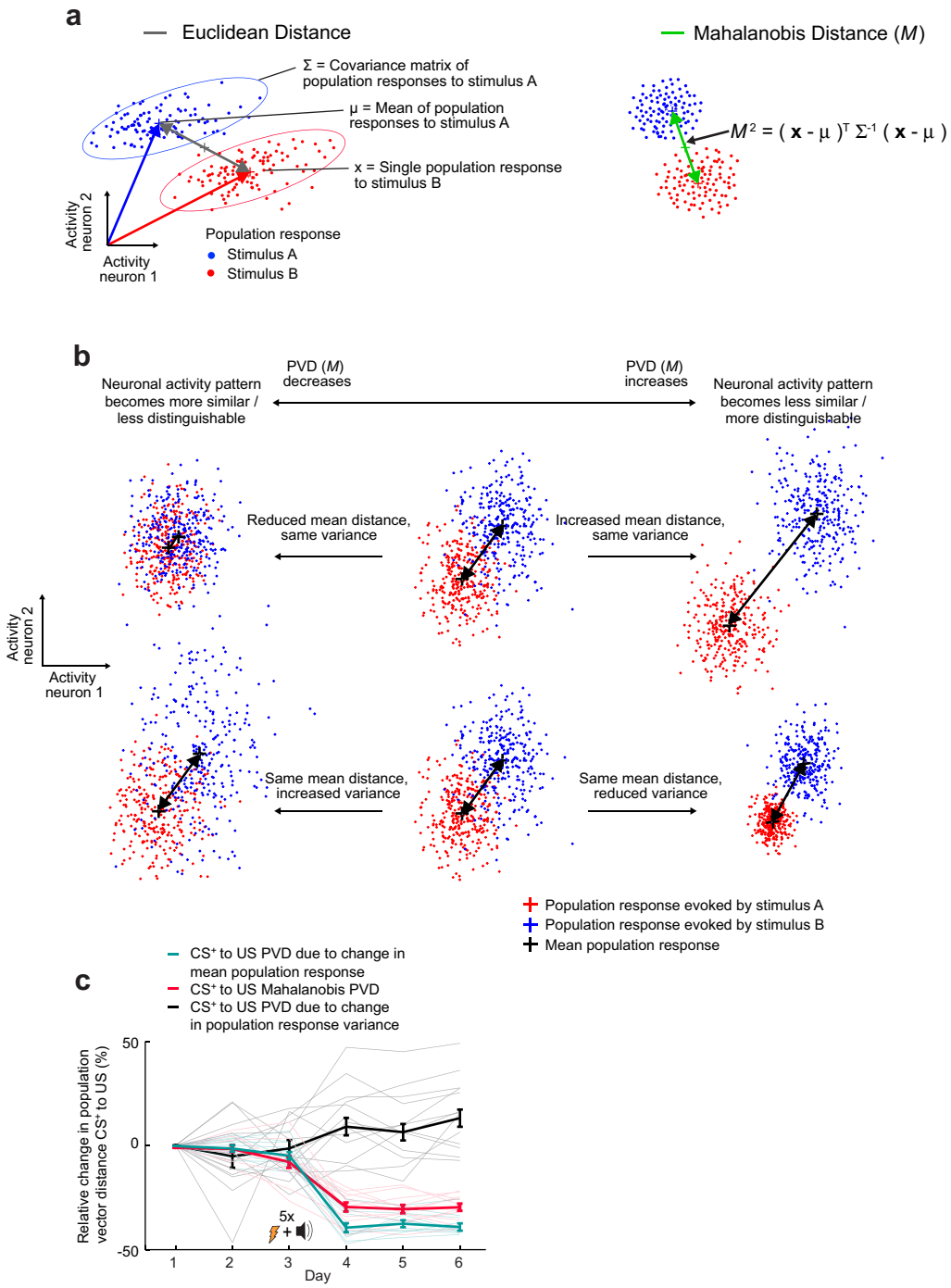
Extended Data Figure 6 | Conditioning induces bi-directional changes in CS-evoked responses. a, **b**, Contrary to the predictions of the cellular, Hebbian model of fear learning, conditioning induced substantial bi-directional changes in the CS^+-evoked responses of cells that responded and those that did not respond to the US. Notably, a preponderance of cells that responded to both the CS^+ and US before training had decreased CS^+-evoked responses after training (**a**). Furthermore, many cells with potentiated CS^+-evoked responses after training were not US-responsive (**b**). Ensemble level analyses showed that cells with up- and downregulated responses made equally important contributions to the learning-induced changes in ensemble neural coding (Fig. 3e). **a**, Percentages of CS^--responsive cells that were also US-responsive (blue) and of CS^+-responsive cells that were also US-responsive. The latter are further divided into cells that increased their CS^+-evoked responses after training (maroon), those that underwent no significant changes in their CS^+-evoked responses (pink), and those that decreased their CS^+-evoked responses after training (red). **b**, Percentages of CS^--responsive cells that were not US-responsive (blue) and of CS^+-responsive cells that were not US-responsive. The latter are further divided into cells that increased their CS^+-evoked responses after training (maroon), those that underwent no significant changes in their CS^+-evoked responses (pink), and those that decreased their CS^+-evoked responses after training (red). All data are from the same 12 mice and denote mean ± s.e.m.



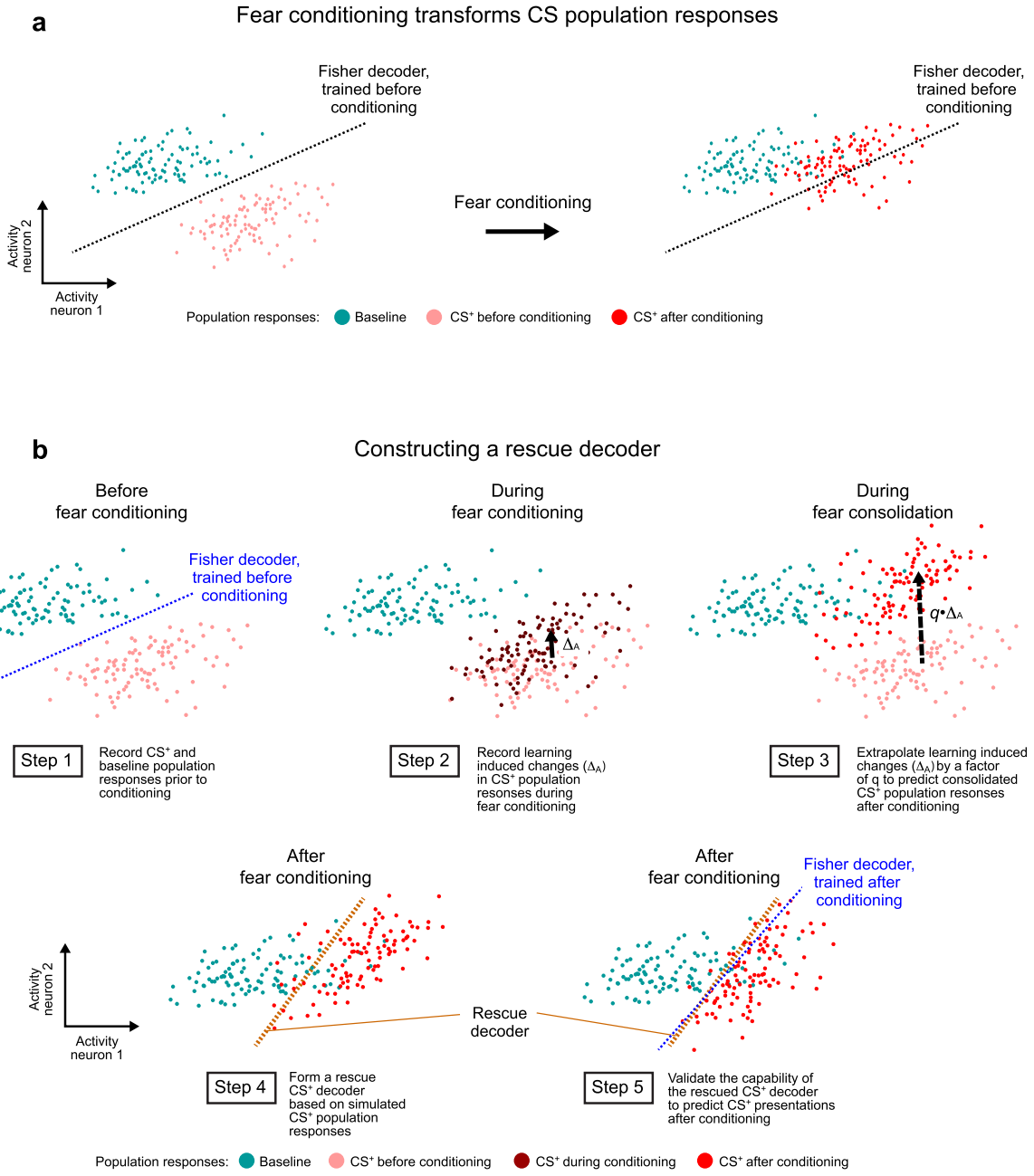
Extended Data Figure 7 | BLA ensembles provide sufficient information to decode the CS, and the decoding accuracy improves with successive tone presentations in a series of tones. a, Left, a three-way decoder has three possible outputs (CS^+, CS^- and baseline) and hence different categories of possible errors. When a decoder makes a type A error, it outputs the wrong CS (for example, CS^+ instead of CS^-). When a decoder makes a type B error, it fails to distinguish a CS presentation from baseline activity. Right, when the activity traces of all neurons were used to train the decoders, the decoders yielded the correct answer on $97 \pm 1\%$ (mean $\pm$ s.e.m.) of all trials from a testing set comprising equal numbers of samples of each type. The success rate was $90 \pm 2\%$ when the decoders were trained using only those cells with statistically significant responses to at least one of the CS types. **b**, For trials that were incorrectly decoded, the pie charts show the proportions of the two types of error, for decoders trained on the activity traces of all neurons (left), and using only neurons with statistically significant responses to at least one of the two CS types (right). **c**, Type A errors declined sharply during the first 5 of the 25 CS tone pulses (black), both before (left), and after (right) conditioning. After conditioning, as the 25 tone pulses proceeded, the mice increasingly distinguished between the CS^- (blue) and the CS^+ (red), as seen by the differences in evoked freezing behaviour. Data are mean $\pm$ s.e.m. **d**, Schematic showing how we extracted the principal components (PCs) of the BLA ensemble responses to CS tone presentations. Dashed box encloses two PCs, used in **e** for illustration. **e**, Plots of the first two PCs, determined as in **d** for four example mice, illustrate that the ensemble responses to the CS^+ (red) and the CS^- (blue) were generally distinguishable. Black stars mark the first out of 25 tone pulses for each CS presentation and illustrate that the initial tones in the series were generally the hardest to categorize correctly.



Extended Data Figure 8 | Fisher linear discriminant analysis of BLA population activity. a, Notional schematic showing the Mahalanobis distance and the discrimination boundary of a Fisher linear discriminant analysis (FLDA) decoder (black dotted line), which discriminates the multi-dimensional, neural ensemble responses to CS presentations from the activity patterns during baseline conditions. For simplicity, the schematic shows a hypothetical case in which the ensemble consisted of only two neurons, but the basic principles readily apply to larger ensembles. For a given set of training data, the Fisher decoder provides the a posteriori probability that a representative data sample will be correctly categorized. **b**, Example histogram from one mouse showing BLA ensemble responses (day 1) to CS presentations, normalized and projected onto the dimension of maximal discriminability. Dashed vertical line marks the classification boundary of the Fisher linear decoder, dividing those ensemble responses classified as baseline from those representing a CS. The separation between the two peaks in the histogram is an empirical estimate of the Mahalanobis distance, which is a multi-dimensional generalization of the discriminability index, d', used in statistics[21]. **c**, Mean decoding performance as a function of the number of cells used for training the decoder (open circles) and corresponding parametric fits to a sigmoid function. When the training and testing data came from the same day (black curve), performance asymptotically approached near perfect decoding when more than ~100 cells were used. When the training and testing data came from different days (red curve), our datasets were not large enough to approach the asymptote. However, the sigmoidal fit suggests that the asymptotic performance of time-lapse decoders would be around 90% in cases with more than 120 cells. Shading indicates s.e.m. **d**, Decoding performance of time-lapse CS^- decoders, as a function of the elapsed time between the day on which the training dataset was acquired and the day on which the testing dataset was acquired ($n = 12$ mice). Despite declining re-occurrence probabilities of the cells as a function of elapsed time (Extended Data Fig. 5f), decoding performance remained stable for time-lapse intervals of 1–5 days. Data are mean ± s.e.m.



Extended Data Figure 9 | The Mahalanobis distance quantifies the discriminability of two sets of ensemble responses and takes into account the mean and covariance of each response set. a, Left, schematic of two sets of ensemble neural responses (blue and red clouds of data points), illustrated for a hypothetical case in which the ensemble consists of only two neurons. The Euclidean distance (grey line) between the means of the two distributions does not take into account the degree to which the ensemble neural responses are variable from trial to trial. Right, to characterize the differentiability of the two response sets in a way that takes into account neural variability, the Mahalanobis distance (M) between the two distributions was determined. To do this, the covariance matrix of the ensemble neural responses (Σ) was used to map the data points into a space in which the distributions have unity variance in all directions. M is equivalent to the distance between the means of the two resulting distributions. **b**, A change in the Mahalanobis PVD can be due to changes in the means of the two sets of ensemble responses, changes in response variability, or both. Schematic illustrates these two different ways in which the sets of ensemble responses can become more or less differentiable. Top row shows a pair of cases in which changes in the mean ensemble responses dominate the change in the PVD. Bottom row shows a pair of cases in which changes in response variability dominate the change in the PVD. **c**, The total change in the CS^+–US PVD (red curve) induced by learning was divided into two components: a component due to changes in the mean CS^+-evoked response (cyan curve) and a component due to changes in the variability of the CS^+-evoked responses (black curve). After conditioning, the CS^+-evoked responses became less variable (black curve) but also more similar to the US-evoked ensemble responses (cyan curve). The latter effect substantially outweighed the former, leading to a net ~32% decline (red curve) in the differentiability of the CS^+- and US-evoked responses, as quantified by the net decrease in the Mahalanobis distance. Thin lines show the values from each of 12 individual mice. Thick lines show the mean values. Error bars are s.e.m.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Procedure for computational rescue of the CS$^+$ decoders. Unlike time-lapse CS$^-$ decoders, which worked well across all 6 days of the experiment, time-lapse CS$^+$ decoders did not work well when the training and testing datasets came from a pair of days that spanned across the conditioning session (Fig. 3b). This failure mode for the CS$^+$ decoders arises from the learning-induced changes in the ensemble representation of the CS$^+$ (Fig. 3c, d). However, by extrapolating the changes in the CS$^+$ representation that occur during the training session on day 3, we could predict the much greater, subsequent changes in the CS$^+$ representation that occur before the next session on day 4 and thereby rescue the failures of the time-lapse CS$^+$ decoders. This figure schematizes the procedure for the computational rescue. **a**, Schematic illustration of how conditioning-induced changes in CS$^+$-evoked ensemble activity (light and dark red dots) can impair the performance of a time-lapse decoder trained on data from before fear conditioning and tested on data from after conditioning. **b**, Using five main steps, we computationally simulated the changes in the CS$^+$-representation that occurred during post-training consolidation, by extrapolating by a factor, q, the much smaller changes in the CS$^+$-representation that occurred during the day 3 training session. **c**, To determine the optimal value of the extrapolation factor q, we simulated the post-training changes in the CS$^+$-representation by computationally adjusting the CS$^+$ population vectors in increments of Δ_A, the modest change in coding that occurred on day 3. Increments of 3–5 times Δ_A were optimal, in that they best rescued the capabilities of two-way decoders trained on either of days 1 or 2 to detect a CS$^+$ presentation when tested on data from after training (days 4–6), or vice versa. (Each datum shows the mean ± s.e.m. decoding performance, averaged across 12 mice and the 12 possible pair-wise combinations per mouse of one pre- and one post-training day.) The same analysis of the CS$^-$ representation scarcely yielded any change in decoding performance, because the effects of training (Δ_A) for the CS$^-$ were negligible. Decoders trained on temporally shuffled data (1,000 shuffles; grey curve) and those based only on cells with up- (green) or downregulated (purple) responses to the CS$^+$ after training performed less successfully than decoders based on all cells (brown).