








# Central amygdala micro-circuits mediate fear extinction

Nigel Whittle<sup>1,2</sup>, Jonathan Fadok <sup>1,6</sup>, Kathryn P. MacPherson<sup>3</sup>, Robin Nguyen<sup>4</sup>, Paolo Botta<sup>1,7</sup>, Steffen B. E. Wolff<sup>1,8</sup>, Christian Müller<sup>1</sup>, Cyril Herry <sup>1,9</sup>, Philip Tovote <sup>1,10</sup>, Andrew Holmes <sup>3</sup>, Nicolas Singewald <sup>2</sup>, Andreas Lüthi <sup>1,5</sup>✉ & Stéphane Cioocchi <sup>4</sup>✉

Fear extinction is an adaptive process whereby defensive responses are attenuated following repeated experience of prior fear-related stimuli without harm. The formation of extinction memories involves interactions between various corticolimbic structures, resulting in reduced central amygdala (CEA) output. Recent studies show, however, the CEA is not merely an output relay of fear responses but contains multiple neuronal subpopulations that interact to calibrate levels of fear responding. Here, by integrating behavioural, in vivo electrophysiological, anatomical and optogenetic approaches in mice we demonstrate that fear extinction produces reversible, stimulus- and context-specific changes in neuronal responses to conditioned stimuli in functionally and genetically defined cell types in the lateral (CEl) and medial (CEm) CEA. Moreover, we show these alterations are absent when extinction is deficient and that selective silencing of protein kinase C delta-expressing (PKC $\delta$ ) CEI neurons impairs fear extinction. Our findings identify CEA inhibitory microcircuits that act as critical elements within the brain networks mediating fear extinction.

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. <sup>2</sup>Department of Pharmacology and Toxicology, Institute of Pharmacy and CMBI, University of Innsbruck, Innsbruck, Austria. <sup>3</sup>Laboratory of Behavioral and Genomic Neuroscience, National Institute on Alcohol Abuse and Alcoholism, NIH, Bethesda, MD, USA. <sup>4</sup>Laboratory of Systems Neuroscience, Department of Physiology, University of Bern, Bern, Switzerland. <sup>5</sup>University of Basel, Basel, Switzerland. <sup>6</sup>Present address: Department of Psychology and Tulane Brain Institute, Tulane University, New Orleans, LA, USA. <sup>7</sup>Present address: Zuckerman Institute, Columbia University, New York, NY, USA. <sup>8</sup>Present address: Department of Pharmacology, University of Maryland School of Medicine, Baltimore, MD, USA. <sup>9</sup>Present address: INSERM, Neurocentre Magendie, U1215, Bordeaux, France. <sup>10</sup>Present address: Institute of Clinical Neurobiology, University Hospital Würzburg, Würzburg, Germany. ✉email: [andreas.luthi@fmi.ch](mailto:andreas.luthi@fmi.ch); [stephane.cioocchi@pyl.unibe.ch](mailto:stephane.cioocchi@pyl.unibe.ch)

The survival of animals depends on their ability to mobilize appropriate defensive behaviours to imminent threats<sup>1</sup>. Yet, animals also need to be able to flexibly adapt to changes in threat contingencies by inhibiting fear responses when threat-related stimuli no longer associate with aversive outcomes, in a process called fear extinction<sup>2–4</sup>. In a typical experimental procedure, an association between conditioned stimuli (CS, a tone or light) and unconditioned stimuli (US, e.g. a foot shock) is first formed (fear conditioning) and then subsequently updated by repeated presentations of the CS in the absence of the US (fear extinction). Fear extinction is thought of as a new learning process wherein animals learn that the CS is no longer predictive of the US<sup>4</sup>. Thus, fear extinction does not reflect the mere erasure of the conditioned fear memory; indeed fear can be spontaneously recovered after fear extinction, or triggered by exposure to the US or a new context<sup>5–7</sup>.

Current models posit that neural circuits and cell assemblies for fear and extinction memories compete with one another, in a context-dependent manner, to determine the degree of fear responding to the CS<sup>4,8,9</sup>. In recent years, there have been major advances in delineating the neural circuitry underlying fear conditioning and extinction<sup>10,11</sup>, yet key elements of this circuitry remain to be elucidated. Of particular note, the central nucleus of the amygdala (CEA) has long been ascribed an essential role in the expression of conditioned fear responses<sup>1</sup>. However, detailed dissection of CEA circuitry, using *in vivo* and *ex vivo* recordings from functionally and/or genetically defined cell types, has recently challenged the view that the CEA is merely an output relay of fear responses. Instead, this work indicates that the CEA contains multiple anatomically, molecularly and functionally defined neuronal subpopulations that interact to calibrate levels of fear responding<sup>12–14</sup>.

CEA output neurons located in both the lateral (CEL) and the medial (CEM) subdivision of the CEA project to downstream targets that mediate different components of conditioned defensive behaviours, such as freezing or flight<sup>15–17</sup>. Fear conditioning potentiates excitatory input onto CEL projections to the ventrolateral periaqueductal grey (vlPAG)<sup>18</sup> and CEM output neurons exhibit increased CS responses upon fear conditioning<sup>15</sup>. In turn, the activity of CEM output neurons is thought to be controlled by excitatory glutamatergic afferents from auditory thalamus and the basal amygdala (BA)<sup>19</sup>.

Crucially, however, CEM output neurons are also subject to inhibitory control from GABAergic neurons located in the neighbouring intercalated cell clusters (ITCs) and a subset of neurons in the CEL expressing protein-kinase C delta (PKC $\delta$ )<sup>15,20,21</sup>. Thus, following fear conditioning, three functional neuronal subpopulations emerge in the CEL: (1) CS ‘non-responsive’ neurons, (2) CELon neurons that are excited by the CS following fear conditioning and overlap in part with a somatostatin (SST)-expressing population<sup>18,22,23</sup> and (3) CELoff neurons, which acquire an inhibitory response to the CS and partly overlap with PKC $\delta$  neurons<sup>15</sup>. Interestingly, CELon and CELoff neurons can inhibit each other<sup>15</sup>, and SST neurons can inhibit PKC $\delta$  neurons<sup>22</sup>, which could result in a switch-like disinhibition of output neurons in CEM or in CEL<sup>12</sup>.

Together, these earlier findings show that there is a layer of processing and plasticity within the CEA that can serve to promote and limit fear responding<sup>24,25</sup>. This raises the intriguing possibility that the same CEA circuitry could be ideally positioned to mediate fear extinction and act as a substrate for the reductions in fear responding that occur with extinction. The major aim of the current study was to test this hypothesis using a combination of behavioural, *in vivo* electrophysiological, optogenetic and molecular approaches. Our

findings demonstrate that microcircuits within the CEA are crucial for fear extinction.

## Results

**Neuronal correlates of fear extinction in subpopulations of CEA neurons.** To first identify neuronal correlates of fear extinction in CEA circuits, we submitted freely-moving mice ( $n = 27$ , C57BL/6J, hereafter B6) to a fear-conditioning and extinction procedure while chronically recording single-unit activity in CEA (Figs. S1, S2; Supplementary Tables 1–4; see ‘Methods’)<sup>15</sup>. Following fear conditioning, mice exhibited a selective increase in conditioned freezing to the CS that was reversed to pre-conditioning levels by the end of fear extinction learning (Fig. 1a, b; Fig. S3).

Examination of the CS-related activity of CEA single units during testing revealed three subpopulations exhibiting distinct patterns of responding that were reversed from fear conditioning to extinction. Fear-conditioning-induced differential conditioned responses in CEM and CEL neurons, as previously reported<sup>15,17</sup>, such that CEM neurons increased their phasic CS responses (Fig. 1c), while CEL neurons exhibited either an inhibitory response (Fig. 1d) or an increase (Fig. 1e) in conditioned responses, consistent with the activity of CELoff and CELon neurons, respectively (Figs. S4, S5). A larger proportion of CELon neurons (64%, 35 out of 55 neurons) exhibited cue-related responses during habituation, as compared to CELoff (24%, 8 out of 33 neurons) and CEM neurons (27%, 4 out of 15 neurons). Strikingly, fear conditioning-related changes in the CS responses of all three neuronal subpopulations were reversed following fear extinction, i.e., elevated CS-related activity in CEM and CELon neurons was attenuated and the inhibited CS-related activity of CELoff neurons was diminished (Fig. 1c–e). These data suggest that changes in CS-related phasic activity within the CEA inhibitory microcircuits signal the extinction of fear responses.

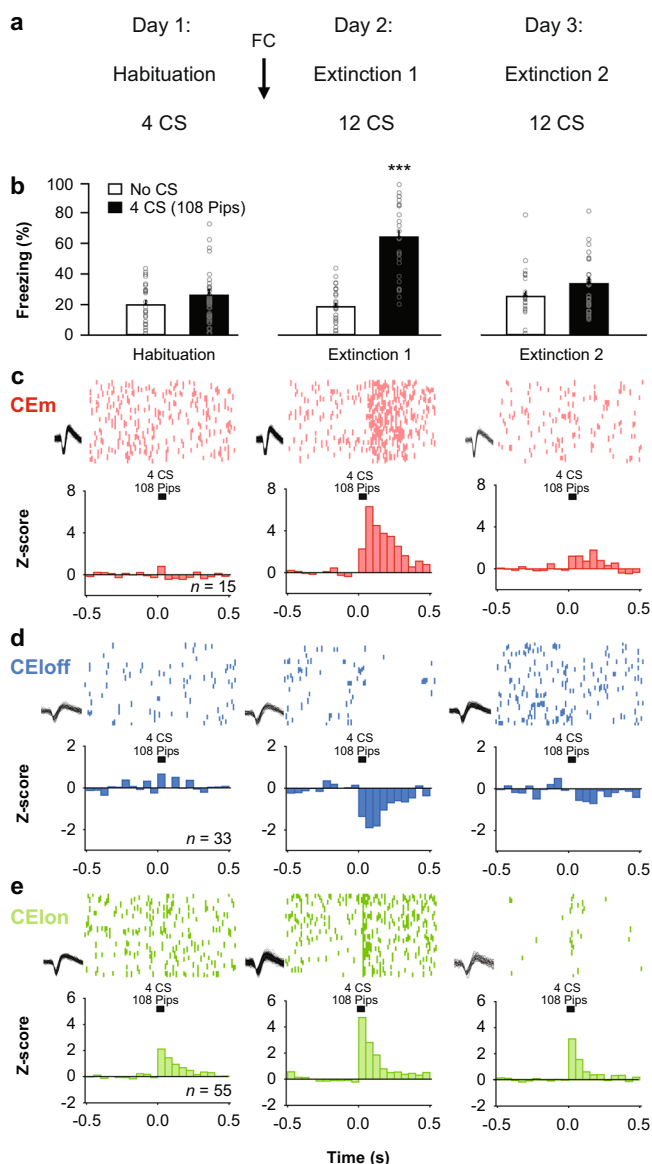
## CEA subpopulation activity tracks extinction-related changes in the expression of fear.

How specific is the reversal of CEA subpopulation activity during fear extinction? Does it correlate with behavioural changes and emotional values of conditioned stimuli, or does it simply reflect a non-associative process such as stimulus habituation or desensitisation? To test for the selectivity of fear extinction-induced reversal of neuronal responses in CEA microcircuits, we trained mice ( $n = 10$ ) using a discriminative fear extinction paradigm (Fig. 2a). In this task, mice were conditioned to two different CSs followed by the extinction of just one of these CSs. Immediately after the extinction, animals were exposed to the non-extinguished CS—which resulted in an instantaneous switch (low to high) in fear behaviour (Fig. 2b).

We reasoned that if the neural responses in CEA subpopulations correlate with fear extinction *per se*, rather than with non-associative or time-related processes, we should expect a change in neural activity paralleling the behavioural switch from the extinguished to the non-extinguished CS. In line with this prediction, we observed that the switch in behaviour corresponded to an immediate recovery of CS-induced responses in CEM, CELoff and CELon neurons to levels of activity evident after fear conditioning (Fig. 2c–e). Thus, CS-related responses in these CEA subpopulations track extinction-induced changes in the expression of fear responses to the CS.

## Context-driven fear renewal reverses extinction-related CEA subpopulation activity.

Fear extinction does not simply reflect forgetting of the CS-US associations, but a new learning process subjected to contextual modulation<sup>4,26</sup>. This sensitivity to the context provided us with an opportunity to test whether changes in CS-related activity in CEA subpopulations is sensitive to contextual changes. One week following fear extinction learning,



**Fig. 1 Neuronal correlates of fear extinction in subpopulations of CEA neurons.** **a** Behavioural protocol. FC: fear conditioning. CS: conditioned stimuli. **b** Behavioural data. B6 mice:  $n = 27$ ; freezing, habituation, no CS:  $19.8 \pm 2.6\%$ , CS:  $26.1 \pm 3.3\%$ , beginning of extinction 1, no CS:  $18.5 \pm 2.3\%$ , CS:  $61.9 \pm 4.6\%$ , end of extinction 2, no CS:  $25.3 \pm 3.1\%$ , CS:  $34.1 \pm 3.4\%$ , blocks (averages) of 4 CSs. One-way repeated-measures ANOVA  $F_{(5,130)} = 30.8$ ,  $p < 0.001$ , followed by post hoc Bonferroni  $t$ -test vs. CS group during habituation,  $p < 0.001$ . Bar plots are expressed as means  $\pm$  SEM. Circles are freezing values of individual mice. **c** Raster plots and corresponding spike waveforms of a representative CEm unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEm neurons:  $n = 15$  units from 5 mice; z-score habituation:  $-0.11 \pm 0.45$ , beginning of extinction 1:  $4.21 \pm 1.75$ , end of extinction 2:  $1.24 \pm 0.48$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(2,28)} = 3.9$ ,  $p = 0.033$  followed by post hoc Bonferroni  $t$ -test vs. during habituation,  $p = 0.023$ . **d** Raster plots and corresponding spike waveforms of a representative CEloff unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEloff neurons:  $n = 33$  units from 18 mice; z-score, habituation:  $0.28 \pm 0.33$ , beginning of extinction 1:  $-1.53 \pm 0.28$ , end of extinction 2:  $-0.46 \pm 0.34$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(2,64)} = 8.4$ ,  $p < 0.001$  followed by post hoc Bonferroni  $t$ -test vs. during habituation,  $p < 0.001$ . **e** Raster plots and corresponding spike waveforms of a representative CELon unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CELon neurons:  $n = 55$  units from 15 mice; z-score, habituation:  $1.30 \pm 0.30$ , beginning of extinction 1:  $2.54 \pm 0.43$ , end of extinction 2:  $1.40 \pm 0.30$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(2,108)} = 5.3$ ,  $p = 0.006$  followed by post hoc Bonferroni  $t$ -test vs. during habituation,  $p = 0.008$ . All individual neurons of each CEA population had significant z-score values upon CS presentation (4 first CSs during extinction 1). Source data are provided as a Source data file.

memory for the extinguished CS was observed in a subset of mice (Fig. 3a, b): these mice expressed low levels of CS-induced freezing when tested in the context in which extinction learning occurred. However, when these same mice were tested in the conditioning context, there was an expected ‘renewal’ of CS-induced conditioned freezing (Fig. 3b, Supplementary Table 5).

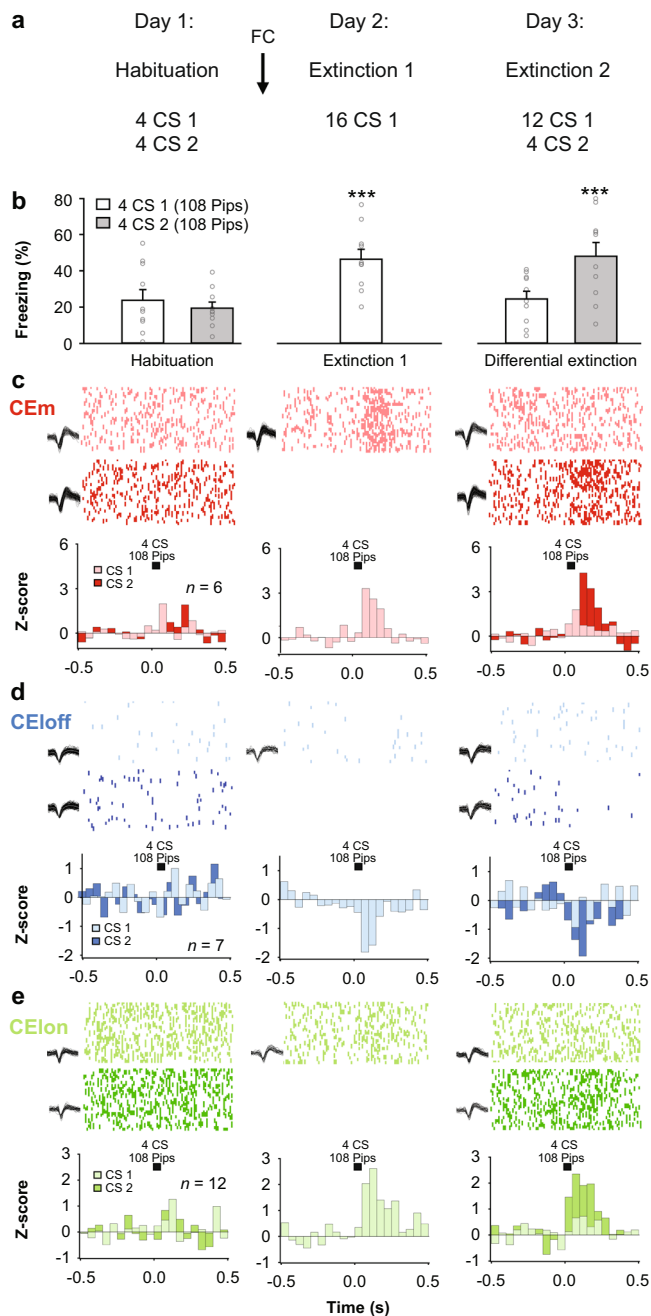
We asked whether this context-induced reversion of fear responding was paralleled by an alteration in CEA single-unit activity. Indeed, while all three CEA neuronal populations showed reduced CS responses during the retrieval of extinction memory in the extinction context (Fig. 3c–e; Fig. S6), CS-related responses were completely re-established to pre-extinction levels when testing occurred in the conditioning context (Fig. 3c–e). Thus, the context-gated expression of extinguished fear memory closely parallels changes in the activity of the CEA subpopulations.

**Impaired extinction associates with persistent fear-related activity in CEA subpopulations.** Impaired fear extinction is a hallmark of anxiety disorders<sup>27</sup>, yet endogenous neural correlates of impaired fear extinction in CEA have not been thoroughly investigated in vivo. Based on our findings thus far, we reasoned

that impaired extinction would correspond with the persistence of a ‘fear-like’ pattern of activity in CEA subpopulations. To test this prediction, we performed CEA single-unit recordings in a mouse strain (129S1/SvImJ, ‘S1’) that exhibits impaired extinction and associated abnormalities in CEA IEG activity<sup>28</sup>. We first replicated prior data<sup>28–30</sup> showing that S1 mice tested using the same procedures used in earlier experiments in the current study, failed to exhibit a reduction in CS-related freezing after extinction training ( $n = 7$  mice, Fig. 4a, b; Fig. S3).

Examination of the CEA single-unit activity in the S1 mice revealed activity patterns in the same three CEA neuronal subpopulations as described above (CELon, CEloff and CEm) (Fig. 1). Strikingly, however, and entirely in keeping with the absence of extinction at the behavioural level, the patterns of CEA activity evident after fear conditioning were largely unchanged after extinction (Fig. 4c–e). Specifically, CEm neurons showed a persistent increase in their phasic CS responses after fear conditioning and extinction, and the inhibitory response of CEloff neurons that emerged after conditioning was also maintained after extinction. Interestingly, however, the conditioning-related increase in the CS-related activity CELon neurons was partially restored to pre-conditioning levels after extinction. This is notable because it shows that CS-related inhibition of CEloff neurons does not require input from CELon neurons and that, by extension, other upstream populations (e.g., ITC neurons) can drive this inhibition.

These data suggest that plastic changes in the activity of CEloff neurons may be a pivotal step in the reduction in fear responses produced by fear extinction, and that the failure of this plasticity may underlie the extinction deficits evident in S1 mice. To further test this hypothesis, we took advantage of the fact that PKC $\delta$  CEL neurons overlap with CEloff



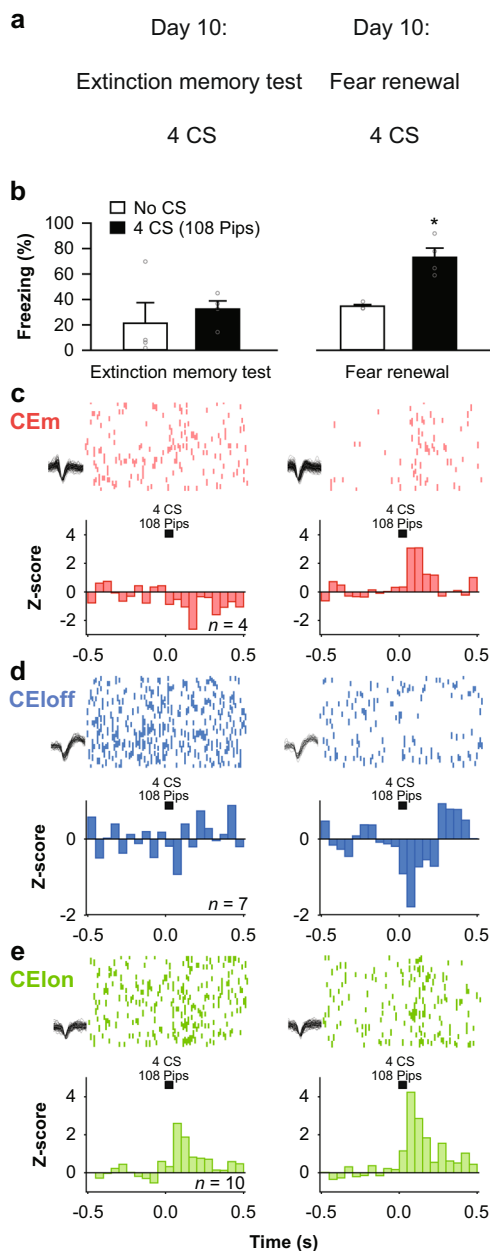
**Fig. 2 CEA subpopulation activity tracks extinction-related changes in the expression of conditioned freezing.** **a** Behavioural protocol. FC: fear conditioning. CS: conditioned stimuli. **b** Freezing behaviour. B6 mice:  $n = 10$ ; freezing, habituation, CS1:  $23.8 \pm 5.9\%$ , CS2:  $19.5.1 \pm 3.3\%$ , beginning of diff. extinction 1, CS1:  $46.4 \pm 5.4\%$ ; end of diff. extinction 2: CS1:  $24.5 \pm 4.3\%$ , CS2:  $48.1 \pm 7.6\%$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(4,36)} = 8.6$ ,  $p < 0.001$ , followed by post hoc Bonferroni  $t$ -test vs. CS2 block during habituation,  $p < 0.001$ . Bar plots are expressed as means  $\pm$  SEM. Circles are freezing values of individual mice. **c** Raster plots and corresponding spike waveforms of a representative CEm unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEm neurons:  $n = 6$  units from 3 mice; z-score, habituation, CS1:  $-0.21 \pm 0.44$ , CS2:  $0.64 \pm 0.38$ , beginning of diff. extinction 1, CS1:  $2.66 \pm 0.87$ ; end of diff. extinction 2: CS1:  $0.70 \pm 0.42$ , CS2:  $3.79 \pm 1.08$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(4,20)} = 5.2$ ,  $p = 0.005$  followed by post hoc Bonferroni  $t$ -test vs. CS1 block during habituation,  $p < 0.05$ . **d** Raster plots and corresponding spike waveforms of a representative CEloff unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEloff neurons:  $n = 7$  units from 5 mice; z-score, habituation, CS1:  $-1.33 \pm 0.47$ ; end of diff. extinction 2: CS1:  $-0.51 \pm 0.72$ , CS2:  $-1.37 \pm 0.35$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(4,24)} = 3.4$ ,  $p = 0.023$  followed by post hoc Bonferroni  $t$ -test vs. CS1 block during habituation,  $p < 0.05$ . **e** Raster plots and corresponding spike waveforms of a representative CElon unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CElon neurons:  $n = 12$  units from 5 mice; z-score, habituation: CS1:  $-0.40 \pm 0.21$ , CS2:  $0.41 \pm 0.23$  beginning of diff. extinction 1: CS1:  $1.60 \pm 0.39$ ; end of diff. extinction 2: CS1:  $0.46 \pm 0.27$ , CS2:  $1.92 \pm 0.53$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(4,44)} = 6.0$ ,  $p < 0.001$  followed by post hoc Bonferroni  $t$ -test vs. CS1 block during habituation,  $p < 0.05$ . All individual neurons of each CEA population had significant z-score values upon CS presentation (first 4 CSs during diff. extinction 1). Source data are provided as a Source data file.

neurons<sup>21</sup> to compare, via immunostaining, the expression of the immediate-early gene (IEG) Zif268 in immunolabeled PKC $\delta$  CEL neurons of S1 and B6 mice after either a fear retrieval or extinction learning test (Fig. S7a). We found that the number of neurons positive for both Zif268 and PKC $\delta$  in CEL was increased in extinguished B6 mice, relative to non-extinguished fear-tested counterparts. By contrast, S1 mice did not show an extinction-related increase in Zif268/PKC $\delta$  neurons, consistent with a failure to reverse CS-related inhibition of the CEloff subpopulation (Fig. S7a). This was not due to a lower number of PKC $\delta$  neurons in CEL of S1 mice, as the overall number of these neurons was similar in the two strains (Fig. S7a). Detailed analysis of the dendritic morphology of CEA neurons in test-naive mice did, however, indicate evidence of more overall dendritic material in CEA neurons of S1, relative to B6 mice, indicating plasticity deficits in CEloff

neurons of S1 mice may relate to underlying structural abnormalities (Fig. S7b).

**Activity of PKC $\delta$ /CEloff neurons during CS exposure is required for extinction memory formation.** Collectively, our findings thus far suggest that extinction-associated neuronal plasticity in CEA circuits may be necessary for the successful acquisition and expression of extinction memories. In particular, our data posit a critical circuit mechanism in which CEloff neurons gate reductions in fear seen with extinction by exerting inhibition of CEm neurons. Alternatively, changes in the activity in these CEA subpopulations during fear extinction could simply reflect the relaying of upstream plasticity mechanisms [e.g., from basolateral amygdala (BLA) or medial prefrontal cortex] to downstream targets (e.g., vPAG). Furthermore, though a prior study found that inactivating CEloff neurons throughout fear conditioning and retrieval increased freezing responses<sup>21</sup>, it remains unclear whether CEloff neurons causally contribute to the decrease in freezing that occurs with extinction.

To address these questions, we took advantage of the fact that PKC $\delta$  CEL neurons overlap with CEloff neurons<sup>21</sup> by using PKC $\delta$ :Cre mice ( $n = 5$ ) to selectively express, in a Cre-dependent manner, the inhibitory opsin, Arch (AAV5-DIO-Arch), in PKC $\delta$  CEA neurons. We then performed optogenetic phototagging experiments to confirm selective control over the activity of these neurons by shining yellow light into the CEA and showing that this reduced the activity of a subset of single units identifiable as PKC $\delta$  neurons (Fig. 5a, b; see 'Methods'). Furthermore, we examined the activity of these photo-identified neurons after fear conditioning and confirmed that the majority (7/12) exhibited an



**Fig. 3 Context-driven fear renewal reverses extinction-related CEA subpopulation activity.** **a** Behavioural protocol. CS: conditioned stimuli. **b** Behavioural data. B6 mice:  $n = 4$ ; extinction memory retrieval, no CS:  $21.4 \pm 16.2\%$ , CS:  $32.5 \pm 6.4\%$ , fear renewal, no CS:  $34.8 \pm 1.2\%$ , CS:  $73.0 \pm 7.3\%$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(3,9)} = 8.5$ ,  $p = 0.005$  followed by post hoc Bonferroni  $t$ -test vs. CS block during extinction memory,  $p = 0.015$ . Bar plots are expressed as means  $\pm$  SEM. Circles are freezing values of individual mice. **c** Raster plots and corresponding spike waveforms of a representative CEm unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEm neurons:  $n = 4$  units from 1 mouse; z-score, extinction memory retrieval:  $-0.79 \pm 1.59$ , fear renewal:  $3.08 \pm 1.06$ , blocks of 4 CSs. Paired student  $t$ -test, two-sided,  $p = 0.039$ . **d** Raster plots and corresponding spike waveforms of a representative CEloff unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEloff neurons:  $n = 7$  units from 2 mice; z-score, extinction memory retrieval:  $-0.04 \pm 0.61$ , fear renewal:  $-0.94 \pm 0.40$ , blocks of 4 CSs. Paired student  $t$ -test, two-sided,  $p = 0.041$ . **e** Raster plots and corresponding spike waveforms of a representative CELon unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CELon neurons:  $n = 10$  units from 2 mice; z-score, extinction memory:  $1.28 \pm 0.41$ , fear renewal:  $2.13 \pm 0.37$ , blocks of 4 CSs. Paired student  $t$ -test, two-sided,  $p = 0.034$ . All individual neurons of each CEA population had significant z-score values upon CS presentation (4 CSs during fear renewal). Source data are provided as a Source data file.

inhibitory response to the CS, consistent with their designation as CEloff neurons (Fig. 5b, Fig. S8).

Next, we tested whether activating or inhibiting the activity of CEL PKC $\delta$  neurons during extinction learning affected freezing responses. To do so, we bilaterally transfected CEL PKC $\delta$  neurons in PKC $\delta$ ::Cre mice ( $n = 11$ ) with an AAV conditionally expressing both the excitatory opsin, channelrhodopsin-2 (ChR2), and the inhibitory opsin, halorhodopsin (eNpHR) (Fig. 5c). Mice were then equivalently fear-conditioned to two CSs (CS1 and CS2), as evidenced by similar freezing responses to both CSs during a fear retrieval test (Fig. 5e, f). On a subsequent extinction training session, during each of the last 12 CS1 presentations, blue light was shone to excite PKC $\delta$  neurons for 300 ms from CS onset—thereby matching the duration of CS-related activity we had observed in the CEloff neurons (see Fig. 1d, Fig. S9). This was followed by 16 presentations of the CS2, during which yellow light was shone over each of the last CS2 to inhibit these same PKC $\delta$  neurons.

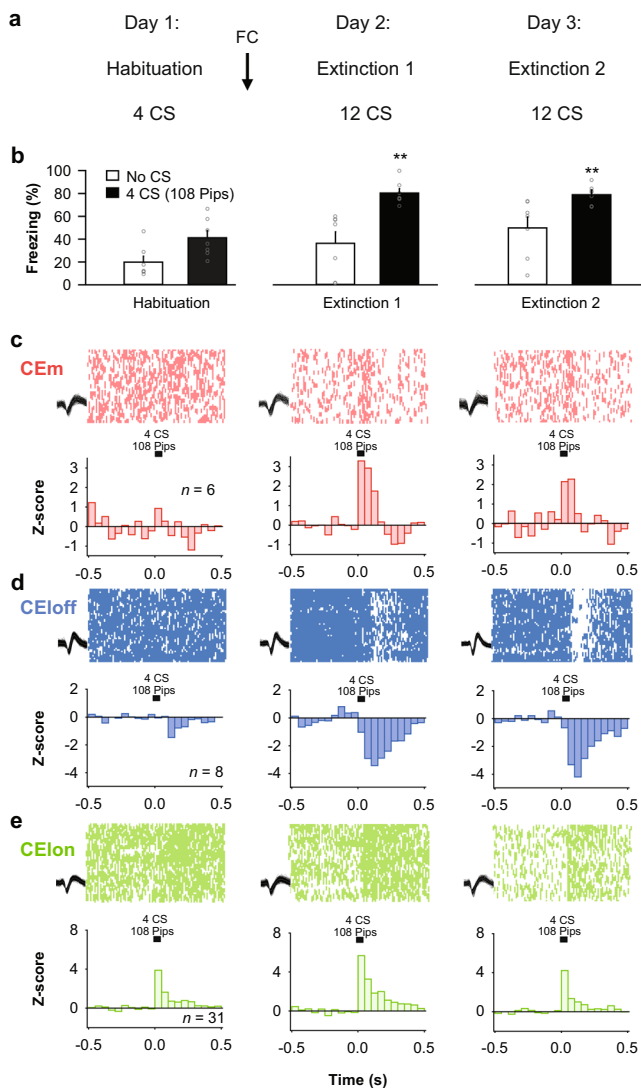
We found that freezing responses did not differ between the CS1/ChR2 and CS2/eNpHR groups during extinction learning, indicating

that manipulating the activity of this subpopulation during a specific temporal window corresponding to CS presentation does not produce acute changes in freezing responses (Fig. 5e). However, when we examined performance during a subsequent extinction memory test, conducted in the absence of light, we found that freezing levels were higher in the CS2/eNpHR than the CS1/ChR2 group (Fig. 5e). In fact, the freezing levels in the CS2/eNpHR group were similar to the levels seen prior to extinction learning, consistent with a failure to form a lasting extinction memory to CS2 when PKC $\delta$ /CEloff neurons were inhibited during extinction training. Importantly, when we repeated the same procedures in PKC $\delta$ ::Cre-negative mice, freezing did not differ between groups at any stage of testing, excluding technical artefacts (Fig. 5f, Fig. S10). Together, these data show that activity of PKC $\delta$ /CEloff neurons during CS exposure is required for the formation of fear extinction memories.

### Discussion

While the CEA is known to play an essential role in the formation and expression of conditioned fear memories, the precise nature of this role is still uncertain<sup>10,11</sup>. Recent studies have shown that fear conditioning potentiates inputs from the lateral amygdala (LA) and paraventricular nucleus of the thalamus (PVT) onto SST-expressing neurons in CEL subdivision of the CEA<sup>14,18</sup>, while inactivation of entire CEL causes fear learning deficits<sup>15</sup>. Furthermore, optogenetically-guided electrophysiological recordings ('phototagging') have demonstrated that SST-expressing CEL neurons correspond to a functional class of 'CELon' neurons<sup>18</sup>, which induce freezing to a CS<sup>13</sup> via direct projections to vPAG<sup>23</sup> or, alternatively, by gating CEm output through disinhibition mediated by PKC $\delta$  ('CEloff') neurons located in CEL<sup>15,21</sup>.

In the current study, we found that CELon neurons exhibited significant cue-related responses during habituation, while these were much less apparent for CEloff and CEm neurons. Hence, CELon neurons might encode an attentional or salience-related signal that could reflect preferential innervation by upstream sensory and attention processing regions. Following fear conditioning, a larger increase in CS-related activity in CELon neurons (e.g., through synaptic plasticity occurring during fear learning at afferent synapses to CELon neurons) may then be



**Fig. 4 Impaired extinction correlates with persistent fear-related activity in CEA subpopulations.** **a** Behavioural protocol. FC: fear conditioning. CS: conditioned stimuli. **b** Behavioural data. S1 mice:  $n = 7$ ; habituation, no CS:  $19.9 \pm 5.1\%$ , CS:  $41.2 \pm 6.3\%$ , beginning of extinction 1, no CS:  $36.4 \pm 10.1\%$ , CS:  $80.3 \pm 3.8\%$ , end of extinction 2, no CS:  $49.8 \pm 9.5\%$ , CS:  $78.9 \pm 3.2\%$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(5,30)} = 12.0$ ,  $p < 0.001$ , followed by post hoc Bonferroni  $t$ -test vs. CS block during habituation,  $p < 0.01$ . Bar plots are expressed as means  $\pm$  SEM. Circles are freezing values of individual mice. **c** Raster plots and corresponding spike waveforms of a representative CEm unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEm neurons:  $n = 6$  units from 2 mice; z-score, habituation:  $0.45 \pm 0.42$ , beginning of extinction 1:  $3.12 \pm 0.73$ , end of extinction 2:  $2.21 \pm 0.48$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(2,10)} = 5.9$ ,  $p = 0.020$  followed by post hoc Bonferroni  $t$ -test vs. during habituation,  $p < 0.05$ . **d** Raster plots and corresponding spike waveforms of a representative CEloff unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CEloff neurons:  $n = 8$  units from 6 mice; z-score, habituation:  $-0.49 \pm 0.87$ , beginning of extinction 1:  $-2.46 \pm 1.13$ , end of extinction 2:  $-2.73 \pm 1.87$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(2,14)} = 4.2$ ,  $p = 0.037$  followed by post hoc Bonferroni  $t$ -test vs. during habituation,  $p < 0.05$ . **e** Raster plots and corresponding spike waveforms of a representative CELon unit (top). Normalized and averaged population peri-stimulus time histograms (bottom). CELon neurons:  $n = 31$  units from 6 mice; z-score, habituation:  $1.60 \pm 0.38$ , beginning of extinction 1:  $3.09 \pm 0.48$ , end of extinction 2:  $1.82 \pm 0.46$ , blocks of 4 CSs. One-way repeated-measures ANOVA  $F_{(2,60)} = 3.8$ ,  $p = 0.028$  followed by post hoc Bonferroni  $t$ -test vs. during habituation,  $p < 0.05$ . All individual neurons of each CEA population had significant z-score values upon CS presentation (first 4 CSs during extinction 1). Source data are provided as a Source data file.

transmitted locally to CEloff neurons, and ultimately disinhibit CEm output neurons to elicit freezing responses.

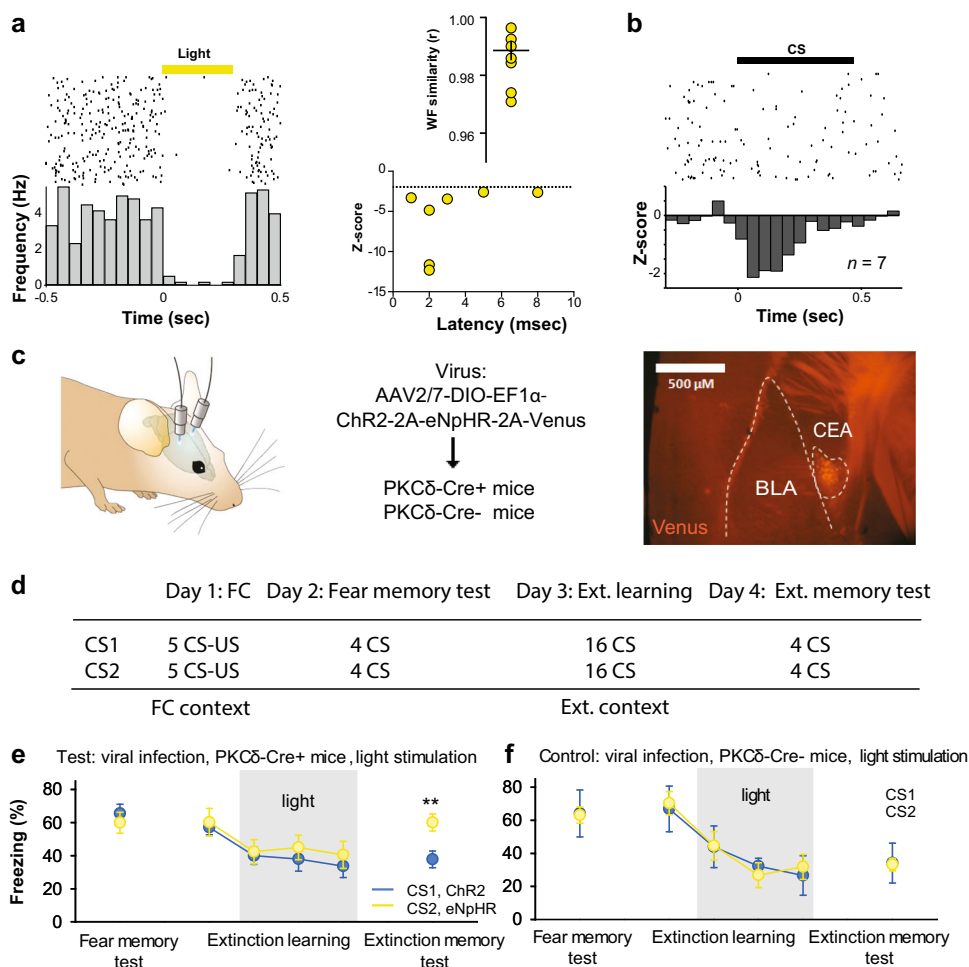
Importantly, we found that the CS responses of CELon neurons were reduced by fear extinction, although some level of CS-related activity, similar to that seen during habituation, remained. This extinction-related reduction was also seen in CELon neurons in mice from the S1 mouse strain, despite these animals showing persistent fear. This implies that persistent fear-related activity in CEloff/CEm neurons is mediated by mechanisms separate from CELon neurons, such as changes in input from neighbouring ITCs. Altogether, these findings provide further evidence that the CEA contains a functionally diverse set of neuronal subpopulations coding for responses to conditioned fear stimulus.

Our findings also demonstrate that stimulus-related responses in these neuronal populations are dynamically modified following extinction to calibrate appropriate levels of freezing. Using a combination of in vivo single-unit recordings and optogenetics, we show that extinction-related reductions in freezing correspond to a relative increase in the CS-related activity of fear-inhibiting CEloff neurons and a decrease in the activity of fear-inducing CELon and CEm neurons. Importantly, we found that these changes in CS-related activity are rapidly reversed when fear responding is renewed by exposure to the fear conditioning context and fail to develop to a non-extinguished CS or in S1 mice that show persistent fear

responding due to impaired extinction. Thus, the activity of these subpopulations of CEA neurons closely tracks shifts in the emotional significance of the CS that occurs with extinction, and is not simply due to non-associative processes, such as habituation or desensitization, that can occur with repeated CS exposure or the passage of time<sup>4,31</sup>. Finally, we identify a key contribution of CEloff neurons in fear extinction by demonstrating that selective photosilencing of PKC $\delta$  CEL neurons during extinction learning prevents extinction memory formation.

PKC $\delta$  neurons in the CEL have been demonstrated to play a critical role in the formation of aversive memories by modulating neuronal activity and plasticity in other brain regions<sup>32,33</sup>. Similarly, our study suggests that PKC $\delta$  neurons may regulate extinction learning by controlling extinction-related synaptic plasticity locally or downstream of the CEL. Thus, PKC $\delta$  neurons may have a more general role in emotional learning by integrating different sensory modalities, valence and attentional signals, thereby flexibly selecting and scaling emotional responses by modulating the activity and plasticity of downstream circuits in motor, autonomic or neuroendocrine centres<sup>12</sup>. Alternatively, distinct subpopulations of PKC $\delta$  neurons might control learning in a valence-specific manner.

The exact circuit and plasticity mechanisms by which CEA neuronal activity is altered by extinction remain to be elucidated. A reduction in the activity of (SST-expressing) CELon occurring with extinction could stem from the reversal (depotential) of the conditioning-induced strengthening of synaptic inputs from LA and PVT onto CEL SST-expressing neurons<sup>14,18</sup>. Alternatively, extinction may engage additional circuit components that suppress the activity of CELon neurons. These components could be extrinsic to the CEL, such as neurons residing in the amygdala striatal transition area<sup>34</sup> or ITC clusters<sup>35</sup>, or intrinsic to the CEL itself, for example in the form of local CEL inhibitory circuits. In



**Fig. 5** PKCδ/CEloff neuronal activity during CS exposure is required for extinction memory formation. **a** Left, PKCδ unit identified with optogenetics. Right, waveform similarity of spikes with or without optogenetic stimulation. Bottom right, latency and magnitude of inhibition of PKCδ/CEloff neurons. **b** Normalized activity (z-score, bottom) of PKCδ/CEloff cells ( $n = 7$ ) and example raster plot (top). CS: conditioned stimuli. **c** An adeno-associated virus (AAV2/7) conditionally expressing Chr2, eNpHR and a Venus reporter under the control of an elongation factor-1α (EF-1α) promoter was injected into the CEI of a PKCδ-Cre+ or PKCδ-Cre- mice (middle). Anti-GFP immunolabelling of the Venus reporter gene in CEI (right). DIO: double-inverted open reading frame; PKCδ: protein-kinase c delta; 2A: ribosomal self-processing peptide; BLA: basolateral amygdala; CEA: central amygdala. **d** Behavioural protocol. **e** Freezing response in PKCδ-Cre+ mice expressing Chr2 and eNpHR in PKCδ neurons. B6 mice:  $n = 11$ ; freezing, Chr2, fear memory, CS I:  $65.5 \pm 5.6\%$ , extinction learning, CS I:  $57.1 \pm 4.5\%$ , CS II:  $39.9 \pm 5.2\%$ , CS III:  $37.9 \pm 7.3\%$ , CS IV:  $33.7 \pm 7.0\%$ , extinction memory, CS I:  $37.7 \pm 5.0$ . eNpHR, fear memory, CS I:  $59.9 \pm 6.3\%$ , extinction learning, CS I:  $60.1 \pm 8.4\%$ , CS II:  $42.4 \pm 7.3\%$ , CS III:  $44.9 \pm 7.4\%$ , CS IV:  $40.5 \pm 8.2\%$ , extinction memory, CS I:  $60.1 \pm 5.2\%$  (blocks of 4 CSs). Main effect of CS presentations during extinction learning: two-way repeated-measures ANOVA  $F_{(5,49)} = 8.7$ ,  $p < 0.001$ . Interaction of light stimulation and CS presentations: two-way repeated-measures ANOVA  $F_{(5,49)} = 2.34$ ,  $p = 0.056$  followed by post hoc Bonferroni  $t$ -test vs. CS1 block during fear memory,  $p < 0.01$ . All values are expressed as means  $\pm$  SEM. **f** Control experiment. Freezing response in PKCδ-Cre- mice not expressing Chr2 and eNpHR in PKCδ neurons. B6 mice:  $n = 5$ ; freezing, blue light, fear memory, CS I:  $63.1 \pm 4.9\%$ , extinction learning, CS I:  $70.4 \pm 9.7\%$ , CS II:  $44.6 \pm 15.0\%$ , CS III:  $26.8 \pm 8.2\%$ , CS IV:  $31.8 \pm 13.5\%$ , extinction memory, CS I:  $33.2 \pm 3.8\%$ . Yellow light, fear memory, CS I:  $64.1 \pm 14.2\%$ , extinction learning, CS I:  $66.9 \pm 12.2\%$ , CS II:  $43.9 \pm 6.1\%$ , CS III:  $32.2 \pm 4.1\%$ , CS IV:  $26.6 \pm 5.7\%$ , extinction memory, CS I:  $34.1 \pm 12.1\%$  (blocks of 4 CSs). Main effect of CS presentations during extinction: two-way repeated-measures ANOVA  $F_{(5,20)} = 15.5$ ,  $p < 0.001$ . Interaction of light stimulation and CS presentations: two-way repeated-measures ANOVA  $F_{(5,20)} = 0.12$ ,  $p = 0.987$  followed by post hoc Bonferroni  $t$ -test vs. CS1 block during fear memory,  $p > 0.05$  for all comparisons. All values are expressed as means  $\pm$  SEM. Source data are provided as a Source data file.

this context, corticotropin-releasing hormone (CRH)-expressing neurons in the CEI have recently been shown to inhibit local SST-expressing neurons, suppress freezing and regulate extinction<sup>22,24,25</sup>. Of further note, CEI CRH-expressing neurons express an array of other neuropeptides and neuropeptide receptors<sup>36</sup>. Given various neuropeptide systems operate in the CEA to control physiological and behavioural readouts of fear<sup>37,38</sup>, defining novel neuropeptide-expressing CEI subpopulations could help identify novel circuit elements underlying extinction-related decreases in CEI neuron activity.

Previous work has shown that the CEI and CEloff populations can modulate one another's activity through reciprocal inhibition<sup>15,21,22</sup>. Hence, a reduction in afferent input from CEI neurons following extinction could produce a release of inhibition over the CEloff subpopulation. Interestingly, our finding that optogenetic silencing of PKCδ (CEloff) neurons impairs extinction memory formation, without producing a frank increase in freezing during silencing, indicates that CEloff neurons do not simply gate fear expression but are a locus of plasticity for extinction. Thus, our study indicates that PKCδ/CEloff

neuronal activity during CS exposure is required for extinction memory formation.

We infer that the diminution of CS-induced inhibitory responses of CEloff neurons over the course of extinction learning and into retrieval (as shown by our electrophysiological data), and associated inhibition of downstream targets, might be necessary for the induction and/or consolidation of a proper extinction memory—for instance by inhibition of freezing-promoting downstream pathways. Optogenetic inhibition of CEloff neurons during CS exposure over the entire extinction acquisition session prevents this shift in CS responses and could therefore impair downstream plasticity events necessary for the formation of a stable long-term extinction memory. Determining the mechanisms underlying changes in the CS response of CEloff neurons during extinction learning will be another important question for future studies. Notwithstanding, our findings are consistent with a heuristic model in which extinction leads to a reduction in the CS-related activity of CELon neurons and a subsequent disinhibition of the CEloff subpopulation, which, in turn, could suppress CEM output to vPAG and thereby reduce freezing.

There are a number of caveats to this model. First, it remains possible that extinction alters the regulation of CEloff neurons by upstream inputs other than (or in addition to) CELon neurons. These inputs could include some of the same aforementioned structures known to innervate CELon cells, such as the PVT, amygdala striatal transition area<sup>34</sup>, BLA and ITC clusters. Second, there is compelling evidence that CEM output can be regulated independently of CEL input. CELon neurons can bypass the CEloff→CEM pathway and project directly to vPAG<sup>18</sup>. Furthermore, extinction is associated with the strengthening of direct, ITC-mediated, feed-forward inhibition of CEM output neurons<sup>39</sup>, in a manner driven by principal cells located in BLA<sup>39</sup> and infralimbic cortex<sup>40</sup>. The observation that permanent ablation of the entire ITC clusters impairs extinction retrieval further highlights a role for these cells<sup>41</sup> though, given evidence of significant heterogeneity between individual ITC clusters<sup>28,30</sup>, the precise nature of this role remains unresolved. Nonetheless, it seems likely that ITCs are a key substrate for extinction that may operate in parallel or in concert with CEloff neurons to affect freezing.

Collectively, these prior observations, taken together with the current findings, suggest that while the CEA is an essential node within the broader neural circuitry mediating fear behaviours, the process of extinction likely engages multiple circuit elements that regulate the activity of CEA neurons to modulate freezing. The combination of independent and interacting circuits would endow a system with the dynamic range and flexibility to adjust behavioural responses to fear-related stimuli according to accumulated experience and prevailing environmental conditions. A system for extinction with flexibility and inbuilt redundancy would be of significant adaptive value considering generating an appropriate level of fear behaviour is crucial to survival for many species. In humans, dysfunction of this system, including deficient plasticity in the CEA subpopulations described here, could contribute to the impaired fear extinction reported in patients with anxiety and trauma-related disorders<sup>42,43</sup>.

## Methods

**Animals.** Male C57BL6/J mice (B6, Harlan Ltd), 129S1/SvImJ mice (S1, Charles River or Jackson Laboratory) and PKC $\delta$ -Cre-CFP mice<sup>21</sup> were housed by strain (1–2 animals per cage) for 7 days before all experiments, under a 12-h light/dark cycle, and were provided with food and water ad libitum. The ambient temperature in the animal facility was ca. 20 °C and the humidity ca. 30%. In the current study, male mice were used to aid comparability with prior analysis of CEA neurons<sup>15,21</sup> and in part because heavier male mice were better suited to carrying the electrode implants during locomotion. It will be important and potentially highly informative to modify procedures to enable the study of fear-related CEA neuronal activity in female mice in future work. All animal procedures were executed in accordance with institutional guidelines and were

approved by the Veterinary Department of the Canton of Basel-Stadt, the Austrian Animal Experimentation Ethics Board and Austrian Ethical Committees on Animal Care and Use (Bundesministerium für Wissenschaft und Verkehr, Kommission für Tierversuchangelegenheiten), the National Institute on Alcohol Abuse and Alcoholism Animal Care and the National Institutes of Health guidelines outlined in 'Using Animals in Intramural Research'.

**Behaviour and optical stimulation.** Fear conditioning and extinction took place in two different contexts (context A and B). The conditioning and extinction boxes and the floor were cleaned with 70% ethanol or 1% acetic acid before and after each session, respectively. To score freezing behaviour, an automatic infrared beam detection system placed on the bottom of the experimental chambers (Coulbourn Instruments) was used. The animals were considered to be freezing if no movement was detected for 2 s. On day 1, mice were submitted to a habituation session in context B, in which they received 4 presentations of the CS (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency: 7.5 kHz or white-noise counterbalanced across animals, 80 dB sound pressure level). Fear conditioning was performed on the same day by pairing the CS with a US (1 s foot shock, 0.6 mA, 5 CS/US pairings; inter-trial interval: 20–180 s). The onset of the US coincided with the offset of the CS. On days 2 and 3, conditioned mice were submitted to extinction training in context B, during which they received 12 presentations of the CS. Retrieval of extinction, spontaneous recovery of conditioned fear (50% freezing cut-off) and context-dependent fear renewal were tested 7 days later in context B and A, respectively, with 4 presentations of the CS<sup>9</sup>. Statistical comparisons were performed with one-way repeated-measures ANOVA followed by Bonferroni post hoc test ( $p < 0.05$  was considered significant).

For the quantification of zif268 in PKC $\delta$  neurons in the CEI amygdala, mice were submitted to an auditory fear conditioning paradigm in which the CS (total CS duration of 30 s, 10 kHz, 80 dB sound pressure level) was paired to the US (2 s foot shock; 0.6 mA; three CS/US pairings; inter-trial interval: 20–180 s) (TSE operant system). The onset of the US coincided with the offset of the CS. Fear conditioning was always performed in a context (context A) different from that used in the extinction session (context B). Context A was cleaned with water and context B with 70% alcohol followed by water. On the following day, fear memory retrieval and extinction training was performed in context B by presenting 16 CSs with an inter-trial interval of 5 s<sup>44</sup>. The 'fear expression' groups received only 2 presentations of the CS following fear conditioning. Freezing behaviour was quantified as an index of fear<sup>45</sup> in each behavioural session by manually quantifying freezing behaviour; defined as no visible movement except that required for respiration, and converted to a percentage [(duration of freezing within the CS/total time of the CS)  $\times$  100] by a trained observer blind to the experimental groups.

For discriminative extinction, mice were habituated on day 1 to 4 presentations of two different CS in context B (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2 ms rise and fall; pip frequency: 7.5 kHz or white noise, 80 dB sound pressure level). Both CSs were subsequently paired with a US (1-s foot shock, 0.6 mA, 5 CS/US pairings for each CS; inter-trial interval: 20–180 s). The onset of the US coincided with the offset of the CS. On days 3 and 4, only one of the two CSs was extinguished by 16 and 12 presentations in context B, respectively. At the end of the second extinction session, mice were exposed to 4 presentations of the non-extinguished CS in context B<sup>9</sup>. Statistical comparisons were performed with one-way repeated-measures ANOVA followed by Bonferroni post hoc test ( $p < 0.05$  was considered significant).

Optogenetic experiments were performed using a fear conditioning and fear extinction procedure in virally injected PKC $\delta$ -Cre positive or negative mice. On day 1, two different CS, CS1 and CS2 (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2 ms rise and fall; pip frequency: 7.5 kHz or white noise, 80 dB sound pressure level, counterbalanced across animals) were paired 5 times with the US (1-s foot shock, 0.6 mA, inter-trial interval: 20–180 s). On day 2, fear memory was tested by presenting 4 CS1 and 4 CS2. On day 3, fear extinction was achieved by sequentially presenting 16 CS1 and 16 CS2 (counterbalanced for order across animals). From the 5th to the 16th CS for CS1 and CS2, each CS pip was coupled to light stimulation (–50 ms to +300 ms from pip onset, 20–40 mW) bilaterally delivered through optic fibres (200  $\mu$ m core diameter, 0.37 NA, Thorlabs GmbH) to the CEI amygdalae.

Optic fibres were connected to a custom-built laser bench using an AOTF (AA Opto-Electronic) to control laser intensity (lasers: MBL473, 473-nm wavelength and MGL593.5, 593.5-nm wavelength, CNILasers). To ensure that animals could move freely, the connecting fibres were suspended over the behavioural context. On day 4, extinction memory was tested by the 4 presentations CS1 and CS2 (counterbalanced for order across animals). Statistical comparisons were performed with two-way repeated-measures ANOVA followed by Bonferroni post hoc test ( $p < 0.05$  was considered significant).

**Single-unit recordings and virus injections.** Mice were anaesthetized with isoflurane (induction 5%, maintenance 2.5%) in O<sub>2</sub>. Body temperature was maintained with a heating pad (CMA/150, CMA/Microdialysis). Mice were secured in a stereotaxic frame and unilaterally implanted in the amygdala with a multi-wire electrode aimed at the following coordinates: 1.3 mm posterior to bregma;  $\pm$ 2.6 mm lateral to midline; 3.25–3.75-mm deep from the cortical surface. The electrodes consisted of 8–16 individually insulated nichrome wires (13  $\mu$ m inner diameter, impedance 50–300 k $\Omega$ ;



California Fine Wire) contained in a 26-gauge stainless steel guide cannula. The wires were attached to a 10 pin to 18 pin connector (Omnetics). The implant was secured using cyanoacrylate adhesive gel. After surgery mice were allowed to recover for 7 days. Analgesia was applied before and during the 3 days after surgery (Metacam).

Electrodes were connected to a headstage (Plexon) containing 8–16 unity-gain operational amplifiers. The headstage was connected to a 16-channel computer-controlled preamplifier (gain X-100, band-pass filter from 150 Hz to 9 kHz, Plexon). Neuronal activity was digitized at 40 kHz and band-pass filtered from 250 Hz to 8 kHz, and was isolated by time–amplitude window discrimination and template matching using a Multichannel Acquisition Processor system (Plexon). At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques<sup>15</sup>.

For optical stimulation of PKC $\delta$  CEI neurons, PKC $\delta$ -Cre+ animals were bilaterally injected into CEI amygdalae with an rAAV serotype 2/7 (Vector Core, University of Pennsylvania), containing a construct conditionally coding for ChR2-2A-eNpHR-2A-Venus<sup>46</sup> at  $-1.3$  mm posterior and  $+/- 2.6$  mm lateral to bregma at a depth of 3.25–3.75 mm. The use of a 2A-Peptide Self-Processing cassette in the AAV2/7 DIO-EF-1 $\alpha$ -ChR2-2A-eNpHR-2A-Venus enables equimolar/isostoichiometric expression of ChR2, eNpHR and Venus in PKC $\delta$  neurons to bi-directionally control their activity<sup>42,47</sup>. For identification of the injection site, the virus solution was mixed at 1:1000 with blue fluorescing polymer microspheres (Duke Scientific Corp.). Deeply anesthetized animals were fixed in a stereotaxic frame (Kopf Instruments) and the skin above the skull was cut. Glass pipettes (tip diameter 10–20  $\mu$ m), connected to a Picospritzer III (Parker Hannifin Corporation), were lowered by a Micropositioner (Kopf Instruments) to the depth of 3.75 mm. About 300 nl were pressure injected bilaterally into CEI amygdalae. In the same surgeries 26-gauge stainless steel guide cannulas (Plastics One) were implanted bilaterally along the same track above CEI amygdalae at a depth of  $-3.25$  mm. Guide cannulas were secured using cyanoacrylate adhesive gel (Henkel) and dental cement (Heraeus Dental). To prevent blockage of the cannulas, dummy cannulas (Plastics One) were inserted and fixed. Behavioural experiments were performed after 4 weeks of recovery and expression time and 3 days of handling. After the experiment, optic fibres were removed and animals were perfused for histological analysis of the injection site as described below.

**Single-unit spike sorting and analysis.** Single-unit spike sorting was performed using Off-Line Spike Sorter (Plexon) as described<sup>15</sup>. Principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces, and clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to be generated from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units and if it displayed a clear refractory period ( $>1$  ms) in the auto-correlogram histograms. In addition, two parameters were used to quantify the overall separation between identified clusters in a particular channel. These parameters include the J3 statistic, which corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies–Bouldin validity index (DB), which reflects the ratio of the sum of within-cluster scatter to between-cluster separation. High values for the J3 and low values for the DB are indicative of good cluster separation. Control values for this statistic were obtained by artificially defining two clusters from the centred cloud of points in the principal component space from channels in which no units could be detected (Supplementary Fig. 1).

Template waveforms were then calculated for well-separated clusters and stored for further analysis. Clusters of identified neurons were analysed offline for each recording session using principal component analysis and a template-matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioural training were considered. Long-term single-unit stability isolation was evaluated using Wavetracker (Plexon) in which principal component space-cylinders were calculated from data recorded during behavioural sessions. Straight cylinders suggest that the same set of single units was recorded during the entire training session (Supplementary Fig. 1). We further quantified the similarity of waveform shape by calculating linear correlation ( $r$ ) values between average waveforms obtained over training days (Supplementary Fig. 1). As a control, we computed the  $r$  values from average waveforms of different neurons.

To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that the reference neuron fires, only one of the two neurons was considered for further analysis. CS-induced neural activity was calculated by comparing the firing rate after stimulus onset with the firing rate recorded during the 500 ms before stimulus onset (bin size, 50 ms; averaged over blocks of 4 CS presentations consisting of 108 individual sound pips in total) using a  $z$ -score transformation.  $Z$ -score values were calculated by subtracting the average baseline firing rate established over the 500 ms preceding stimulus onset from individual raw values and by dividing the difference by the baseline standard deviation. Classification of units was performed by considering a significant  $z$ -score value within 250 ms after CS onset during the fear test according to the freezing levels. Normalized populations PSTHs were obtained by averaging normalized PSTHs from individual neurons. Statistical comparisons were performed with one-way repeated-measures ANOVA followed by Bonferroni post hoc test or with the Student paired  $t$ -test for the recall and renewal datasets ( $p < 0.05$  was considered

significant). Calculations were made in MATLAB and R. Statistical analysis was done in the commercially available software GraphPad Prism and SigmaPlot.

**Optical identification of single units.** For optogenetic identification of PKC $\delta$  neurons, we used pulses of yellow light (to activate Arch). We used 300-ms pulses, 120 times, with a 2 s inter-pulse interval, at 10 mW light power at the fibre tip. Units were considered as light responsive if they showed significant, time-locked ( $<10$  ms) changes in neuronal activity upon illumination. To determine the onset of inhibition, we used change-point analysis (Change Point Analyzer 2.0, Taylor Enterprises Inc.). As described previously<sup>22,46</sup>, this identifies the time point exhibiting a significant change in neuronal activity relative to the preceding time points. We calculated linear correlation ( $r$ ) values for spontaneous and light-evoked spikes to quantitatively determine the similarity of their waveform shapes.

**Immunohistochemistry and imaging.** After completion of experiments, virally injected PKC $\delta$ -Cre+ mice were deeply anaesthetized with avertin (0.3 g/kg). Mice were then transcardially perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (PFA). Coronal, 80- $\mu$ m-thick brain slices were then cut with a vibratome (VT1000 S, Leica) and stored in PBS containing 0.05% sodium azide. To visualize virus expression, standard immunolabelling procedures were performed on free-floating brain sections: overnight incubation at 4 °C with goat rabbit anti-GFP antibody (1:1000, catalogue no. A11122, Invitrogen), 2 h incubation with anti-rabbit Alexa 488 (1:1000, catalogue no. A11008, Invitrogen). After a final wash, slices were mounted on cover slips and imaged. Mice were included in the analysis if they showed virus expression bilaterally within CEI amygdalae and if fibre tip placement was not more than  $\sim 500$   $\mu$ m away from CEI amygdala.

For the quantification of zif268 in PKC $\delta$  neurons in the CEI amygdala, mice were killed 2 h after the start of the extinction training session as previously described<sup>30</sup>. Mice were deeply anesthetized using sodium pentobarbitone (200 mg/kg) and transcardially perfused with 20 ml of 0.9% saline followed by 20 ml of 4% paraformaldehyde in phosphate-buffered saline (PBS), pH 7.4. Samples were post-fixed for 2 h in the same fixative at 4 °C and stored in PBS. Coronal sections (40  $\mu$ m) were cut on a vibratome (Leica Microsystems) and collected in tris buffered saline (TBS). Free-floating sections were incubated in blocking solution (10% BSA and 0.1% Triton X-100 in TBS) then with primary polyclonal rabbit anti-Zif268 antibody (1:2000; Cat. No.: sc-189; Santa Cruz Biotechnology) and monoclonal mouse anti-PKC $\delta$  (1:1000, Cat. No.: 610398, BD Transduction Laboratories) for 48 h at 4 °C. The sections were then washed with TBS and incubated for 2 h at room temperature with Cy2-conjugated donkey anti-rabbit (1:500; Cat. No.: 711-225-152; Jackson ImmunoResearch Laboratories) and Alexa Fluor 647-conjugated donkey anti-mouse (1:500; Cat. No.: 717-605-150; Jackson ImmunoResearch Laboratories).

Sections were then attached to microscope slides and coverslipped with FluoroGold Antifade reagent. All immunolabelled sections were imaged using an Olympus BX51 microscope equipped with an Olympus XM10 video camera. Images taken under consistent exposure times using a  $\times 20$  oil-immersed optical objective lens (UPlanSApo, Olympus Corporation) were digitised and viewed using CellSens Dimension 1.5 software (Olympus Corporation, Tokyo, Japan). The quantification of Zif268 expression in PKC $\delta$  positive or negative expressing neurons in the CEI was achieved by manual scoring. Statistical comparisons were performed with a two-way ANOVA followed by Fischer LSD post hoc test ( $p < 0.05$  was considered significant).

**Dendritic morphology of CEI neurons.** The dendritic morphology of CEI neurons was determined using Golgi stain as described previously<sup>29,48</sup>. Mice were overdosed with xylazine/ketamine and then transcardially perfused with 0.9% saline. Brains were removed and immersed in Golgi-Cox solution (1:1 solution of 5% potassium dichromate and 5% mercuric chloride diluted 4:10 with 5% potassium chromate) for 18 days. Brains were dehydrated, infiltrated with a graded series of celloidins, and embedded in 8% celloidin. Coronal sections were cut at a thickness of 160  $\mu$ m on a sliding microtome (American Optical 860) and alkalized, developed, fixed, dehydrated, cleared, mounted, and coverslipped.

Neurons selected for reconstruction did not have truncated branches and were unobscured by neighbouring neurons and glia, with dendrites that were easily discriminable by focusing through the depth of the tissue. In 4–6 sections evenly spaced through the rostral-caudal extent of the CEI amygdala, an average of 4–6 neurons per mouse (average of 2.5 from each hemisphere) were randomly selected (using a random number generator, <http://www.randomizer.org>) and reconstructed. Neurons were drawn in three-dimensions by an experimenter blind to strain, using a  $\times 100$  objective on an Olympus BX41 system microscope using a computer-based neuron tracing system (NeuroLucida, MBF Biosciences). The length and number of dendrites, as well as the length and number of terminal branches, was measured for all dendritic arbours. Values were compared between strains using  $t$ -tests. In addition, to assess the overall amount and location of dendritic material, a three-dimensional version of a Sholl analysis<sup>49</sup> was performed by measuring the number of dendritic intersections within 10  $\mu$ m concentric spheres radiating from the soma.

**Statistics and reproducibility.** Imaging was repeated independently with similar results in Fig. 5c ( $n = 11$  mice) and in Supplementary Fig. 7a (B6 expression,  $n = 8$  mice; S1 expression,  $n = 6$  mice; B6 extinction,  $n = 8$  mice; S1 extinction,  $n = 6$  mice). Distinct spike waveforms recorded from different units and sorted using 3D principal component analysis could be observed in all recordings with more than one unit per electrode as shown in Supplementary Fig. 1b.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The data that support the findings on this study are available from the corresponding authors upon request. The data that supports the findings of this study are available at <https://data.fmi.ch/PublicationSupplementRepo/>. Source data are provided with this paper.

### Code availability

The computer code that supports the findings of this study is available from the corresponding authors upon request.

Received: 2 July 2020; Accepted: 28 May 2021;

Published online: 06 July 2021

### References

- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Maren, S. Seeking a spotless mind: extinction, deconsolidation, and erasure of fear memory. *Neuron* **70**, 830–845 (2011).
- Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* **12**, 120–150 (2007).
- Herry, C. Neuronal circuits of fear extinction. *Eur. J. Neurosci.* **31**, 599–612 (2010).
- Bouton, M. Context and ambiguity in the extinction of emotional learning: implications for exposure therapy. *Behav. Res. Ther.* **26**, 137–149 (1988).
- Bouton, M. E. Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biol. Psychiatry* **52**, 976–986 (2002).
- Corcoran, K. A. & Maren, S. Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *J. Neurosci.* **21**, 1720–1726 (2001).
- Zelikowsky, M., Pham, D. L. & Fanselow, M. S. Temporal factors control hippocampal contributions to fear renewal after extinction. *Hippocampus* **22**, 1096–1106 (2011).
- Herry, C. Switching on and off fear by distinct neuronal circuits. *Nature* **454**, 600–606 (2008).
- Tovote, P., Fadok, J. P. & Luthi, A. Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
- Bukalo, O., Pinard, C. R. & Holmes, A. Mechanisms to medicines: elucidating neural and molecular substrates of fear extinction to identify novel treatments for anxiety disorders. *Br. J. Pharm.* **171**, 4690–4718 (2014).
- Fadok, J. P., Markovic, M., Tovote, P. & Luthi, A. New perspectives on central amygdala function. *Curr. Opin. Neurobiol.* **49**, 141–147 (2018).
- Yu, K., Garcia da Silva, P., Albeanu, D. F. & Li, B. Central amygdala somatostatin neurons gate passive and active defensive behaviors. *J. Neurosci.* **36**, 6488–6496 (2016).
- Penzo, M. A. et al. The paraventricular thalamus controls a central amygdala fear circuit. *Nature* **519**, 455–459 (2015).
- Ciocchi, S. Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468**, 277–282 (2010).
- Tovote, P. et al. Midbrain circuits for defensive behaviour. *Nature* **534**, 206–212 (2016).
- Duvarci, S., Popa, D. & Pare, D. Central amygdala activity during fear conditioning. *J. Neurosci.* **31**, 289–294 (2011).
- Penzo, M. A., Robert, V. & Li, B. Fear conditioning potentiates synaptic transmission onto long-range projection neurons in the lateral subdivision of central amygdala. *J. Neurosci.* **34**, 2432–2437 (2014).
- Keifer, O. P. Jr., Hurt, R. C., Ressler, K. J. & Marvar, P. J. The physiology of fear: reconceptualizing the role of the central amygdala in fear learning. *Physiology* **30**, 389–401 (2015).
- Paré, D., Quirk, G. J. & LeDoux, J. E. New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* **92**, 1–9 (2004).
- Haubensak, W. Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature* **468**, 270–276 (2010).
- Fadok, J. P. et al. A competitive inhibitory circuit for selection of active and passive fear responses. *Nature* **542**, 96–100 (2017).
- Li, H. et al. Experience-dependent modification of a central amygdala fear circuit. *Nat. Neurosci.* **16**, 332–339 (2013).
- Jo, Y. S., Namboodiri, V. M. K., Stuber, G. D. & Zweifel, L. S. Persistent activation of central amygdala CRF neurons helps drive the immediate fear extinction deficit. *Nat. Commun.* **11**, 422 (2020).
- Hartley, N. D. et al. Dynamic remodeling of a basolateral-to-central amygdala glutamatergic circuit across fear states. *Nat. Neurosci.* **22**, 2000–2012 (2019).
- Bouton, M. E., Westbrook, R. F., Corcoran, K. A. & Maren, S. Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biol. Psychiatry* **60**, 352–360 (2006).
- Singewald, N. & Holmes, A. Rodent models of impaired fear extinction. *Psychopharmacology* **236**, 21–32 (2019).
- Hefner, K. et al. Impaired fear extinction learning and cortico-amygdala circuit abnormalities in a common genetic mouse strain. *J. Neurosci.* **28**, 8074–8085 (2008).
- Camp, M. C. et al. Genetic strain differences in learned fear inhibition associated with variation in neuroendocrine, autonomic, and amygdala dendritic phenotypes. *Neuropsychopharmacology* **37**, 1534–1547 (2012).
- Whittle, N., Hauschild, M., Lubeck, G., Holmes, A. & Singewald, N. Rescue of impaired fear extinction and normalization of cortico-amygdala circuit dysfunction in a genetic mouse model by dietary zinc restriction. *J. Neurosci.* **30**, 13586–13596 (2010).
- Kamprath, K. & Wotjak, C. T. Nonassociative learning processes determine expression and extinction of conditioned fear in mice. *Learn. Mem.* **11**, 770–786 (2004).
- Yu, K. et al. The central amygdala controls learning in the lateral amygdala. *Nat. Neurosci.* **20**, 1680–1685 (2017).
- Cui, Y. et al. A central amygdala-substantia innominata neural circuitry encodes aversive reinforcement signals. *Cell Rep.* **21**, 1770–1782 (2017).
- Xu, C. et al. Distinct hippocampal pathways mediate dissociable roles of context in memory retrieval. *Cell* **167**, 961–972 e16 (2016).
- Paré, D. & Smith, Y. The intercalated cell masses project to the central and medial nuclei of the amygdala in cats. *Neuroscience* **57**, 1077–1090 (1993).
- Babaev, O., Piletti Chatain, C. & Krueger-Burg, D. Inhibition in the amygdala anxiety circuitry. *Exp. Mol. Med.* **50**, 18 (2018).
- Huber, D., Veinante, P. & Stoop, R. Vasopressin and oxytocin excite distinct neuronal populations in the central amygdala. *Science* **308**, 245–248 (2005).
- Triana-Del Rio, R., van den Burg, E., Stoop, R. & Hegoburu, C. Acute and long-lasting effects of oxytocin in cortico-limbic circuits: consequences for fear recall and extinction. *Psychopharmacology* **236**, 339–354 (2019).
- Amano, T., Unal, C. T. & Pare, D. Synaptic correlates of fear extinction in the amygdala. *Nat. Neurosci.* **13**, 489–494 (2010).
- Berretta, S., Pantazopoulos, H., Caldera, M., Pantazopoulos, P. & Paré, D. Infralimbic cortex activation increases c-Fos expression in intercalated neurons of the amygdala. *Neuroscience* **132**, 943–953 (2005).
- Likhtik, E., Popa, D., Apergis-Schoute, J., Fidacaro, G. A. & Pare, D. Amygdala intercalated neurons are required for expression of fear extinction. *Nature* **454**, 642–645 (2008).
- Botta, P. et al. Regulating anxiety with extrasynaptic inhibition. *Nat. Neurosci.* **18**, 1493–1500 (2015).
- Roberto, M., Kirson, D. Khom, S. The role of the central amygdala in alcohol dependence. *Cold Spring Harb. Perspect. Med.* **11**, a039339 (2021).
- Whittle, N. et al. Deep brain stimulation, histone deacetylase inhibitors and glutamatergic drugs rescue resistance to fear extinction in a genetic mouse model. *Neuropharmacology* **64**, 414–423 (2013).
- Blanchard, R. J. & Blanchard, D. C. Crouching as an index of fear. *J. Comp. Physiol. Psychol.* **67**, 370–375 (1969).
- Wolff, S. B. E. et al. Amygdala interneuron subtypes control fear learning through disinhibition. *Nature* **509**, 453 (2014).
- Senn, V. et al. Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron* **81**, 428–437 (2014).
- Izquierdo, A., Wellman, C. L. & Holmes, A. Brief uncontrollable stress causes dendritic retraction in infralimbic cortex and resistance to fear extinction in mice. *J. Neurosci.* **26**, 5733–5738 (2006).
- Wellman, C. L. et al. Impaired stress-coping and fear extinction and abnormal corticolimbic morphology in serotonin transporter knock-out mice. *J. Neurosci.* **27**, 684–691 (2007).

### Acknowledgements

This work was supported by the following grants to the following individuals: the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 669582), and a Swiss National Science Foundation core grant (310030B\_170268) all to A.L., an ERC starting grant (716761) and a Swiss National Science Foundation professorship grant (170654) (S.C.), by grants from the Austrian Science Fund (FWF) (P25375-B24, I2433-B26) (to N.S. and N.W.), the

NIAAA Intramural Research Program (A.H. and K.P.M.), a Schering Foundation fellowship (S.B.E.W.), and by the Novartis Research Foundation.

### Author contributions

N.W., J.F., K.P.M., R.N., P.B., S.B.E.W., C.M., C.H., P.T., A.H., N.S., A.L. and S.C. contributed to experiments, data analysis, and preparation of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24068-x>.

**Correspondence** and requests for materials should be addressed to A.Lüt. or S.C.

**Peer review information** *Nature Communications* thanks Anna Beyeler, Mario Penzo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021