Creating a False Memory in the Hippocampus

Creating a False Memory in the Hippocampus

Steve Ramirez,[1]† Xu Liu,[1,2]† Pei-Ann Lin,[1] Junghyup Suh,[1] Michele Pignatelli,[1]
Roger L. Redondo,[1,2] Tomás J. Ryan,[1,2] Susumu Tonegawa[1,2]*

[1]RIKEN-MIT Center for Neural Circuit Genetics at the Picower Institute for Learning and
Memory, Department of Biology and Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
[2]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA
02139, U.S.A.

† These authors contributed equally to this work.
* To whom correspondence should be addressed: tonegawa@mit.edu

One Sentence Summary:
We successfully encoded a false association between a contextual memory and an aversive
stimulus by substituting presentation of a natural conditioned stimulus with optical reactivation
of contextual memory engram-bearing dentate gyrus cells.

Memories can be unreliable. Here, we created a false memory in mice by optogenetically manipulating memory engram-bearing cells in the hippocampus. Dentate gyrus (DG) or CA1 neurons activated by exposure to a particular context were labeled with channelrhodopsin-2 (ChR2). These neurons were later optically reactivated during fear conditioning in a different context. The DG experimental group showed increased freezing in the original context in which a foot shock was never delivered. The recall of this false memory was context-specific, activated similar downstream regions engaged during natural fear memory recall, and was also capable of driving an active fear response. Our data demonstrate that it is possible to generate an internally represented and behaviorally expressed fear memory via artificial means.

Neuroscience aims to explain how brain activity drives cognition. Doing so requires identification of the brain regions that are specifically involved in producing internal mental representations and perturbing their activity to see how various cognitive processes are affected. More specifically, humans have a rich repertoire of mental representations generated internally by processes such as conscious or unconscious recall, dreaming, and imagination (*1*, *2*). However, whether these internal representations can be combined with external stimuli to generate new memories has not been vigorously studied.

Damage to the hippocampus impairs episodic memory (*3–8*). Recently, using fear conditioning in mice as a model of episodic memory, we identified a small subpopulation of granule cells in the DG of the hippocampus as contextual memory-engram cells. Optogenetic stimulation of these cells is sufficient to activate behavioral recall of a context-dependent fear memory formed by a delivery of footshocks. This finding provided an opportunity to investigate how the internal representation of a specific context can be associated with external stimuli of high valence. In particular, a hypothesis of great interest is whether artificially activating a previously formed contextual memory engram while simultaneously delivering footshocks can result in the creation of a false fear memory for the context in which footshocks were never delivered. To address this, we investigated whether a light-activated contextual memory in the DG or CA1 can serve as a functional conditioned stimulus (CS) in fear conditioning.

Our system utilizes c-fos-tTA transgenic mice in which the promoter of the *c-fos* gene drives the expression of the tetracycline transactivator (tTA) to induce expression of a gene of interest downstream of the tetracycline-responsive element (TRE) (*8–12*). We injected an AAV

virus encoding TRE-ChR2-mCherry into the DG or CA1 of c-fos-tTA animals (Fig. 1A). ChR2-mCherry expression was completely absent in the DG of animals that had been raised with doxycycline (Dox) in the diet (on Dox) (Fig. 1B). Exploration of a novel context under the condition of Dox withdrawal (off Dox) elicited an increase in ChR2-mCherry expression (Fig. 1C). We confirmed the functionality of the expressed ChR2-mCherry by recording light-induced spikes in cells expressing ChR2-mCherry from both acute hippocampal slices and in anaesthetized animals (Fig. 1D–F). Furthermore, optical stimulation of ChR2-mCherry-expressing DG cells induced cFos expression throughout the anterior-posterior axis of the DG (fig. S1A–I).

We first took virus-infected and fiber-implanted animals off Dox to open a time window for labeling cells activated by the exploration of a novel context (context A) with ChR2-mCherry. The animals were then put back on Dox to prevent any further labeling. The next day, we fear-conditioned this group in a distinct context (context B) while optically reactivating the cells labeled in context A. On the following two days, we tested the animals' fear memory in either the original context A or a novel context C (Fig. 1G). If the light-reactivated cells labeled in context A can produce a functional conditioned stimulus (CS) during fear conditioning in context B, then the animals should express a false fear memory by freezing in context A, but not in context C.

First we examined the degree of overlap of the cell populations activated in contexts A and C (*8,11*). We injected a group of c-fos-tTA mice with an AAV virus encoding TRE-ChR2-mCherry and exposed them to context A while off Dox to label activated DG cells with ChR2-

mCherry. These animals were then immediately placed back on Dox to prevent further labeling. The next day, half of the animals were exposed to context C and the other half were re-exposed to context A as a control. Both groups were sacrificed 1.5 hours later. DG cells activated by the first exposure to context A were identified by ChR2-mCherry expression and cells activated by the exposure to context C or the re-exposure to context A were identified by the expression of endogenous c-Fos. The c-Fos generated by the first exposure to context A had been degraded by the time the animals underwent their second context exposure (*11*). Contexts A and C recruited statistically independent populations of DG cells. In contrast, two exposures to context A recruited substantially overlapping cell populations in the dorsal DG (Fig. 2A–E).

When DG cells activated by the exposure to context A were reactivated with light during fear conditioning in a distinct context B, the animals subsequently froze in context A at levels significantly higher than the background levels, while freezing in context C did not differ from background levels (Fig. 2F). This increased freezing in context A was not due to generalization because a control group expressing only mCherry that underwent the exact same training protocol did not show the same effect (Fig. 2F). A separate group of animals expressing ChR2-EYFP instead of ChR2-mCherry in the DG that underwent the same behavioral schedule also showed increased freezing in context A (fig. S2A).

New experimental and control groups of mice were taken off Dox in context A to label activated cells, and then placed in context C on the following day while back on Dox. In this experiment, although conditioning took place after the formation of both context A and context C memories, only those cells encoding context A were reactivated by light during fear

conditioning. Subsequently, all groups of mice displayed background levels of freezing in context C. In contrast, in the context A test the next day, the experimental group showed increased freezing levels compared to the mCherry-only group, confirming that the recall of the false memory is specific to context A (Fig. 2G). This freezing was not observed in another ChR2-mCherry group that underwent the same behavioral protocol but without light stimulation during fear conditioning in context B, or in a group in which an immediate shock protocol was administered in context B with light stimulation of context A cells (Fig. 2G, fig. S3). In a separate group of animals, we labeled cells active in context C rather than context A and repeated similar experiments as above. These animals showed freezing in context C but not context A (fig. S2B).

The hippocampus processes mnemonic information by altering the combined activity of subsets of cells within defined subregions in response to discrete episodes (*11–13*). Therefore, we investigated whether applying the same parameters and manipulations to CA1 as we did to the DG could form a false memory. We first confirmed that light could activate cells expressing ChR2-mCherry along the anterior-posterior axis of the CA1 similar to the DG (fig. S1J–R). Also similar to the DG (Fig. 2A–E), the overlap of active CA1 cells was significantly lower across contexts (A and C) compared to a re-exposure to the same context (A and A). However, the degree of overlap for the two contexts was much greater in CA1 (30%) than in the DG (~1%). When we labeled CA1 cells activated in context A and reactivated these cells with light during fear conditioning in context B, no increase in freezing was observed in the experimental group expressing ChR2-mCherry compared to the mCherry-only control group in either context A or context C, regardless of whether the animals were exposed to context C or not prior to fear

conditioning in context B (Fig. 2M, N).

The simultaneous availability of two CS's can sometimes result in competitive conditioning; the memory for each individual CS is acquired less strongly compared to when it is presented alone, and the presentation of two simultaneous CS's to animals trained with a single CS can also lead to decrement in recall (*14*). In our experiments, it is possible that the light-activated DG cells encoding context A interfered with the acquisition or expression of the genuine fear memory for context B. Indeed, upon re-exposure to context B, the experimental group froze significantly less than the group that did not receive light during fear conditioning or the group expressing mCherry alone (Fig. 3A and fig. S4). During light-on epochs in the context B test, freezing increased in the experimental group and decreased in the group that did not receive light during fear conditioning (Fig. 3A, fig. S2C). We conducted similar experiments with mice in which the manipulation was targeted to the CA1 region and found no differences in the experimental or control groups during either light-off or light-on epochs of the context B test (fig. S5).

Memory recall can be induced for a genuine fear memory by light reactivation of the corresponding engram in the DG (*8*). To investigate if this applies to a false fear memory, we examined fear-memory recall of experimental and control groups of mice in a distinct context (context D) with light-off and light-on epochs (Fig. 3B). All groups exhibited background levels of freezing during light-off epochs. The experimental group, however, froze at significantly higher levels (~25%) during light-on epochs. This light-induced freezing in context D was not observed in control animals that underwent the same behavioral schedule but did not receive

light during fear conditioning in context B, in animals expressing mCherry alone, in animals receiving immediate shock, or in animals in which CA1 was manipulated instead (Fig. 3B, figs. S2D, S3C, S4 and S5).

Moreover, we quantified the levels of c-Fos expression in the basolateral amygdala (BLA) and the central amygdala (CeA) during the recall of a false and genuine fear memory (*15–20*). Both sessions elicited a significant increase in c-Fos-positive cells in the BLA and CeA compared to a control group exploring a neutral context (Fig. 3C–F).

Finally, a new cohort of mice was trained in a conditioned place avoidance (CPA) paradigm (*21*). Naïve animals did not show an innate preference for either chamber across multiple days (fig. S6). An experimental group injected with the ChR2-mCherry virus and a control group injected with the mCherry-only virus were taken off Dox and exposed to one chamber of the CPA apparatus to label the DG cells activated in this chamber. These animals were then placed back on Dox and on the following day were exposed to the opposite chamber. Next, the mice were fear conditioned in a different context with light stimulation. The following day, they were placed back into the CPA apparatus and their preference between the chambers was measured (Fig. 4A). After conditioning, the experimental group showed a strong preference for the unlabeled chamber over the labeled chamber, whereas the mCherry-only group spent an equal amount of time exploring both chambers (Fig. 4B–D and fig. S6). Notably, exposure to the two chambers activated statistically independent population of DG cells (Fig. 4E–K). We conducted similar behavioral tests targeting the CA1 subregion of the hippocampus and the experimental group did not show any chamber preference (Fig. 4L, M).

8

Our results show that cells activated previously in the hippocampal DG region can subsequently serve as a functional CS in a fear conditioning paradigm when artificially reactivated during the delivery of a US. The consequence is the formation of a false associative fear memory to the CS that was not naturally available at the time of the US delivery. This is consistent with previous findings that high frequency stimulation of the perforant path, an input to DG, can serve as a CS in a conditioned suppression paradigm (22).

Memory is constructive in nature; the act of recalling a memory renders it labile and highly susceptible to modification (23,24). In humans, memory distortions and illusions occur frequently. These phenomena often result from the incorporation of misinformation into memory from external sources (25–27). Cognitive studies in humans have reported robust activity in the hippocampus during the recall of both false and genuine memories (28). However, human studies utilizing behavioral and fMRI techniques have not been able to delineate the hippocampal subregions and circuits that are responsible for the generated false memories. Our experiments provide an animal model in which false and genuine memories can be investigated at the memory engram level (29). We propose that optical reactivation of cells that were naturally activated during the formation of a contextual memory induced the retrieval of that memory and the retrieved memory became associated with an event of high valence (i.e. a foot shock) to form a new but false memory. Thus, the experimental group of animals showed increased freezing in a context in which they were never shocked (context A). Although our design for the formation and expression of a false memory was for a laboratory setting, and the retrieval of the contextual memory during conditioning occurred by artificial means (i.e. light),

we speculate that the formation of at least some false memories in humans may occur in natural settings by internally driven retrieval of a previously formed memory and its association with concurrent external stimuli of high valence.

Our experiments also allowed us to examine the dynamic interaction between the false and genuine memories at different stages of the memory process. During the acquisition phase, the artificial contextual information (context A by light activation) either competed with the genuine contextual cues (context B by natural exposure) for the valence of the US (foot shock), or may have interfered with the perception of the genuine contextual cues. This resulted in reduced expression of both false and genuine fear memories compared to the strength of recall attainable after normal fear conditioning (Fig. 3A, compare the two groups during the light-off epoch). This could also be related to the overshadowing effects for multiple CS's (*30*). During the recall phase in context B, the false memory and the genuine memory were either additive (Fig. 3A, compare the with-light group during light-off and light-on epochs) or competitive (Fig. 3A, compare the no light group during light-off and light-on epochs). All these observations are consistent with the predictions of an updated Rescorla-Wagner componential model for two independent CS's and suggest that the light-activated artificial CS is qualitatively similar to the genuine CS (*14*).

A previous study applied a similar experimental protocol with pharmacosynthetic methods and failed to see increased freezing upon re-exposure to either context A or context B. Instead, they observed a synthetic memory that could only be retrieved by the combination of both contexts A and B (*9*). A key difference in their system is that the c-Fos-expressing cells in the entire forebrain were labeled and reactivated over an extended period by a synthetic ligand. We

propose that activating neurons in much wider spatial and temporal domains may favor the formation of a synthetic memory, which may not be easily retrievable by the cues associated with each individual memory. In contrast, activating neurons in a more spatially (only small populations of DG cells) and temporally restricted manner (only a few minutes during light stimulation) may favor the formation of two distinct (false and genuine) memories as observed in our case. In line with this hypothesis, when we manipulated CA1 cells by the same procedures as the ones used for DG cells, we could not create a false memory (i.e. freezing in context A). In CA1, the overlap of the cell populations activated by consecutive exposures to a pair of contexts is much greater than in the DG. While additional work is needed to reveal the nature of CA1 engrams, we hypothesize that our negative CA1 behavioral data could be a result of contextual engrams relying less on a population code and increasingly on a temporal code as they travel through the trisynaptic circuit (*4,11−13*).

References and Notes:

1. D.L. Schacter, D.R. Addis, R.L. Buckner, *Nat. Rev. Neurosci.* **8**, 657–661 (2007).

2. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsaki, *Science* **321**, 1322–1327 (2008).

3. H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, I. Fried, *Science* **322**, 96–101 (2008).

4. C.J. MacDonald, K.Q. Lepage, U.T. Eden, H. Eichenbaum, *Neuron* **71**, 737–749 (2011).

5. G. Buzsaki, E.I. Moser, *Nat. Neurosci.* **16**, 130–138 (2013).

6. T.J. McHugh, *Science* **317**, 94–99 (2007).

7. D. Tse *et al.*, *Science* **316**, 76–82 (2007).

8. X. Liu *et al.*, *Nature* **484**, 381–385 (2012).

9. A.R. Garner *et al.*, *Science* **335**, 1513–1516 (2012).

10. L.G. Reijmers, B.L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230–1233 (2007).

11. S. Kubik, T. Miyashita, J.F. Guzowski, *Learn. Mem.* **14,** 758–770 (2007).

12. J.F. Guzowski, B.L. McNaughton, C.A. Barnes, P.F. Worley, *Nat. Neurosci.* **2**, 1120–1124 (1999).

13. J.K. Leutgeb, S. Leutgeb, M.B. Moser, E.I. Moser, *Science* **315,** 961–966 (2007).

14. S.E. Brandon, E.H. Vogel, A.R. Wagner, *Behav. Brain Res.* **110**, 67–72 (2010).

15. J.H. Han *et al.*, *Science* **323**, 1492–1496 (2009).

16. M. Rogan, U. Staubli, J. LeDoux, *Nature* **390**, 604–607 (1997).

17. J.P. Johansen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12692–12697 (2010).

18. S. Maren, G.J. Quirk, *Nat. Rev. Neurosci.* **5,** 844–852 (2004).

19. H. Li *et al.*, *Nat. Neurosci.* **16**, 332–339 (2013).

20. S. Ciocchi *et al.*, *Nature* **468**, 277–282 (2010).

21. S. Lammel *et al.*, *Nature* **491**, 212–217 (2012).

22. V. Doyere, S. Laroche, *Hippocampus* **2**, 39–48 (1992).

23. K. Nader, G.E. Schafe, J.E. LeDoux, *Nature* **406,** 722–726 (2000).

24. F.C. Bartlett, *Remembering: A study in experimental and social psychology* (Cambridge Univ. Press, Cambridge, 1932).

25. E.F. Loftus, *Nat. Rev. Neurosci.* **4**, 231–234 (2003).

26. D.L. Schacter, E.F. Loftus, *Nat. Neurosci.* **16,** 119–123 (2013).

27. H.L. Roediger, K.B. McDermott, *J. Exp. Psychol. Learn.* **24**, 803–814 (1995).

28. R. Cabeza, S.M. Rao, A.D. Wagner, A.R. Mayer, D.L. Schacter, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4805–4810 (2001).

29. S.M. McTighe, R.A. Cowell, B.D. Winters, T.J. Bussey, L.M. Saksida, *Science* **330,** 1408–1410 (2010).

30. I. P. Pavlov, *Conditioned reflexes* (Oxford University Press, Oxford, 1927).

Fig. 1. Activity-dependent labeling and light-activation of hippocampal neurons, and the basic experimental scheme. (**A**) The c-fos-tTA mice were bilaterally injected with AAV$_9$-TRE-ChR2-mCherry and implanted with optical fibers targeting DG. (**B**) While on Dox, exploration of a novel context did not induce expression of ChR2-mCherry. (**C**) While off Dox, exploration of a novel context induced expression of ChR2-mCherry in DG. (**D**) Light pulses induced spikes in a CA1 neuron expressing ChR2-mCherry. The recorded neuron is shown labeled with biocytin in (**E**). (**F**) Light pulses induced spikes in DG neurons recorded from a head-fixed anesthetized c-fos-tTA animal expressing ChR2-mCherry. (**G**) Basic experimental scheme. Post-surgery mice were taken off Dox and allowed to explore context A to let DG or CA1 cells become labeled with ChR2-mCherry. Mice were put back on Dox and fear conditioned in context B with simultaneous delivery of light pulses. Freezing levels were then measured in both the original context A and a novel context C. The light green shading indicates the presence of Dox in the diet during corresponding stages of the scheme. Prime (') indicates the second exposure to a given context. The yellow lightning symbol and blue shower symbol indicate foot shocks and blue light delivery, respectively. Red circles represent neurons encoding context A that are thus labeled with ChR2-mCherry. Gray and white circles represent neurons encoding context B and C, respectively. Asterisks (*) indicate neurons activated either by exposure to context or light stimulation.

Fig. 2. Creation of a false contextual fear memory. (**A–E**) c-fos-tTA mice injected with AAV$_9$-TRE-ChR2-mCherry in the DG were taken off Dox and exposed to context A to label the activated cells with mCherry (red), then put back on Dox and exposed to the same context A (A and E) or a novel context C (B and D) 24 h later to let activated cells express c-Fos (green). Images of the DG from these animals are shown in (A) to (D), and the quantifications are shown in (E) ($n = 4$ subjects each; ***$P < 0.001$, unpaired Student's t-test). Blue and red dashed lines indicate the chance level of overlap for A-A and A-C groups, respectively. (**F**) Top: Training and testing scheme of animals injected with AAV$_9$-TRE-ChR2-mCherry or AAV$_9$-TRE-mCherry. Various symbols are as explained in Fig. 1. Bottom: Animals' freezing levels in context A before fear conditioning and in context A and C after fear conditioning ($n = 8$ for ChR2-mCherry group

and $n = 6$ for mCherry group; ***$P < 0.001$, two-way analysis of variance (ANOVA) with repeated measures followed by Bonferroni post-hoc test). (**G**) Top: Training and testing scheme of animals injected with AAV$_9$-TRE-ChR2-mCherry or AAV$_9$-TRE-mCherry. One control group injected with AAV$_9$-TRE-ChR2-mCherry did not receive light stimulation during fear conditioning (ChR2-mCherry, No light). Bottom: Animals' freezing levels in context A and C before and after fear conditioning ($n = 11$ for ChR2-mCherry group and $n = 12$ for mCherry, and $n = 9$ for ChR2-mCherry, No light groups; ***$P < 0.001$, two-way ANOVA with repeated measures followed by Bonferroni post-hoc test). (**H–L**) Animals underwent the same protocol as in (A) to (E), except the virus injection was targeted to CA1. Representative images of CA1 from these animals are shown in (H) to (K), and the quantifications are shown in (L) ($n = 4$ subjects each; *$P = 0.009$, unpaired Student's t-test). (**M**) Same as (F), except the viral injection and implants were targeted to CA1 ($n = 8$ for ChR2-mCherry and mCherry groups; n.s., not significant, two-way ANOVA with repeated measures followed by Bonferroni post-hoc test). (**N**) Same as (G), except the viral injection and implants were targeted to CA1 ($n = 6$ for ChR2-mCherry group and $n = 5$ for mCherry group). Scale bar in (A) and (H) is 250 μm.

Fig. 3. The false and genuine fear memories interact with each other and both recruit the amygdala. (**A**) Animals that underwent the behavioral protocol shown in Fig. 2G were re-exposed to context B and the freezing levels were examined both in the absence and presence of light stimulation ($n = 11$ for ChR2-mCherry group and $n = 9$ for ChR2-mCherry, No light group; *$P = 0.027$; ***$P < 0.001$; #$P = 0.034$, two-way analysis of variance (ANOVA) with repeated measures followed by Bonferroni post-hoc test). (**B**) Animals that underwent the behavioral protocol shown in (A) were placed in a novel context D and the freezing levels were examined both in the absence and presence of light stimulation ($n = 11$ for ChR2-mCherry group and $n = 9$ for ChR2-mCherry, No light group; **$P = 0.007$, two-way analysis of variance (ANOVA) with repeated measures followed by Bonferroni post-hoc test). (**C**) Three groups of mice underwent the training shown in (A) and were sacrificed after testing in either context B (natural recall), A (false recall), or C (neutral context). The percentage of c-Fos-positive cells was calculated for each group in basolateral amygdala (BLA) and central amygdala (CeA) ($n = 6$ subjects each; ***$P < 0.001$). Images for natural recall, false recall, or neutral context are shown in (**D**), (**E**), and (**F**).

Fig. 4. The false memory supports active fear behavior. (**A**) The scheme for conditioned place avoidance paradigm. Various symbols are as explained in Fig. 1. (**B**) Locomotion traces during testing from animals injected with AAV$_9$-TRE-mCherry (top), or animals injected with AAV$_9$-TRE-ChR2-mCherry and DG cells subsequently labeled, corresponding to either the left (middle) or right (bottom) chamber. (**C** and **D**) ChR2-mCherry and mCherry group preferences for the labeled vs. unlabeled chambers as shown by the ratio (C) or the difference in duration of the time spent in each chamber (D). ($n = 8$; *$P = 0.013$; **$P = 0.008$, unpaired Student's t-test). The red dashed line indicates no preference. (**E**–**K**), c-fos-tTA mice injected with AAV$_9$-TRE-EYFP in the DG were taken off Dox and exposed to one chamber to label the activated cells with EYFP (green), then put back on Dox and exposed to the opposite chamber 24 h later to let activated cells express c-Fos (red). Expression of EYFP (E and H), expression of c-Fos (F and I), and a merged view (G and J) are shown. Filled arrows indicate cells expressing EYFP. Hollow arrows indicate cells expressing c-Fos. These cells appear yellow because they express both endogenous c-Fos (red) and the nuclear-localized c-fos-shEGFP (green) from the mouse line (*10*). Quantifications from the dorsal blades of the DG are shown in (K) ($n = 4$). Red dashed lines indicate the chance level of overlap. (**L** and **M**) Same as (C and D), except the viral injection and implants were targeted to CA1 ($n = 6$ each group).
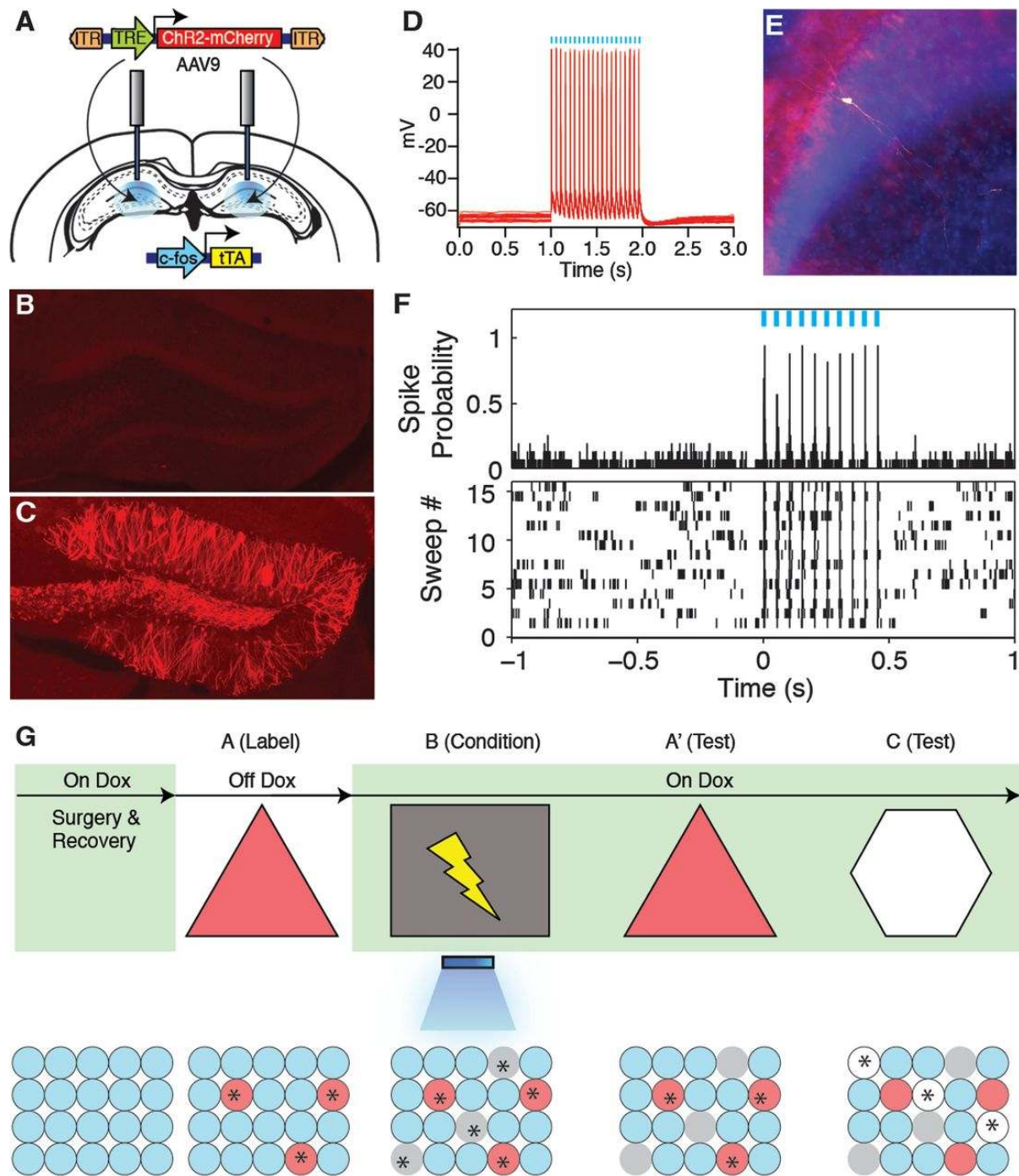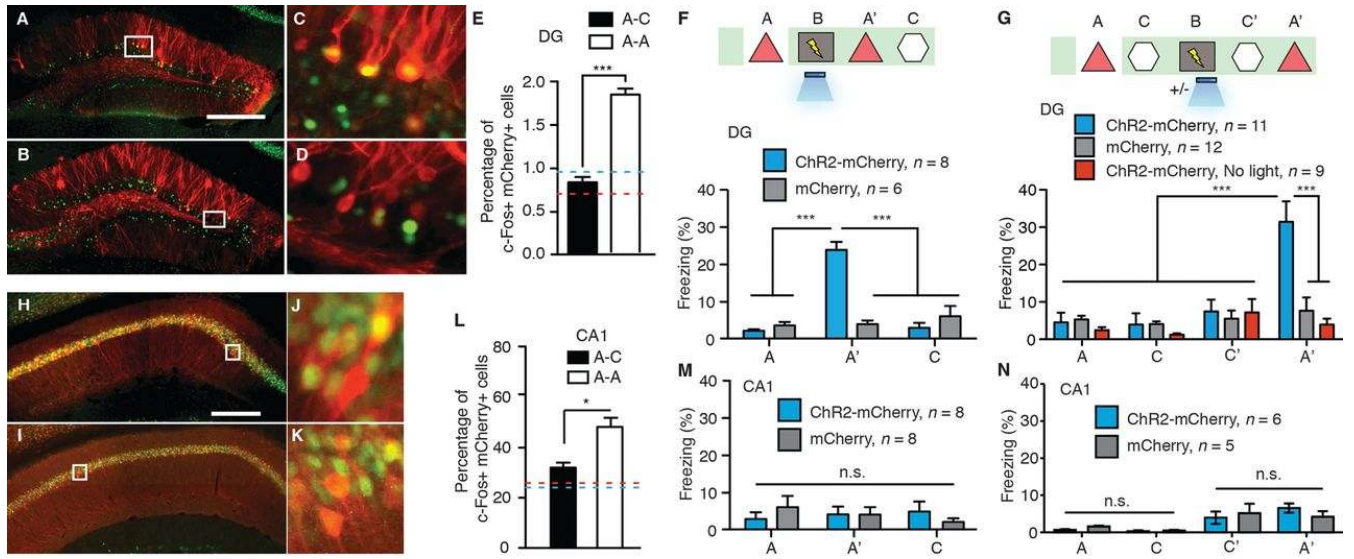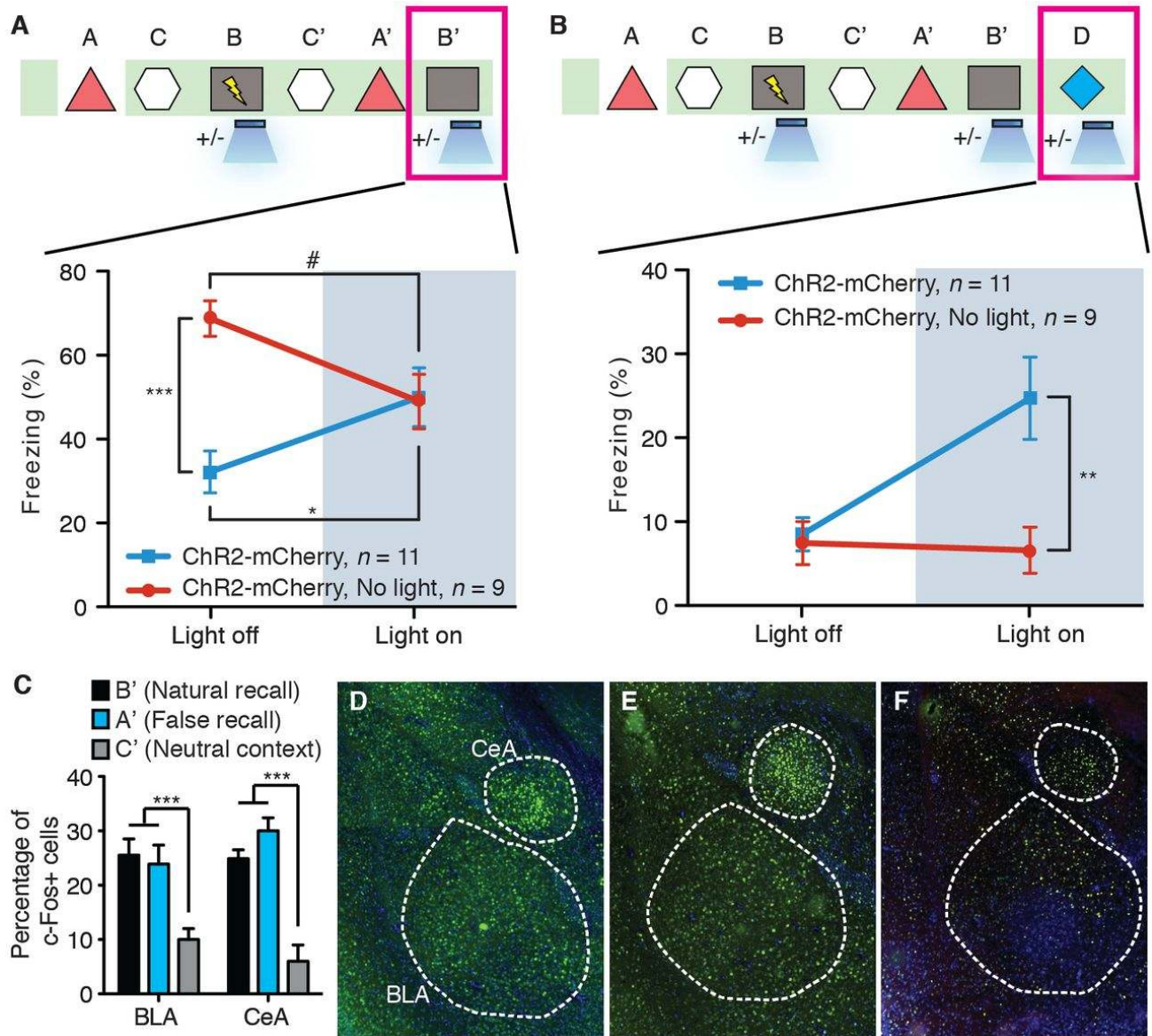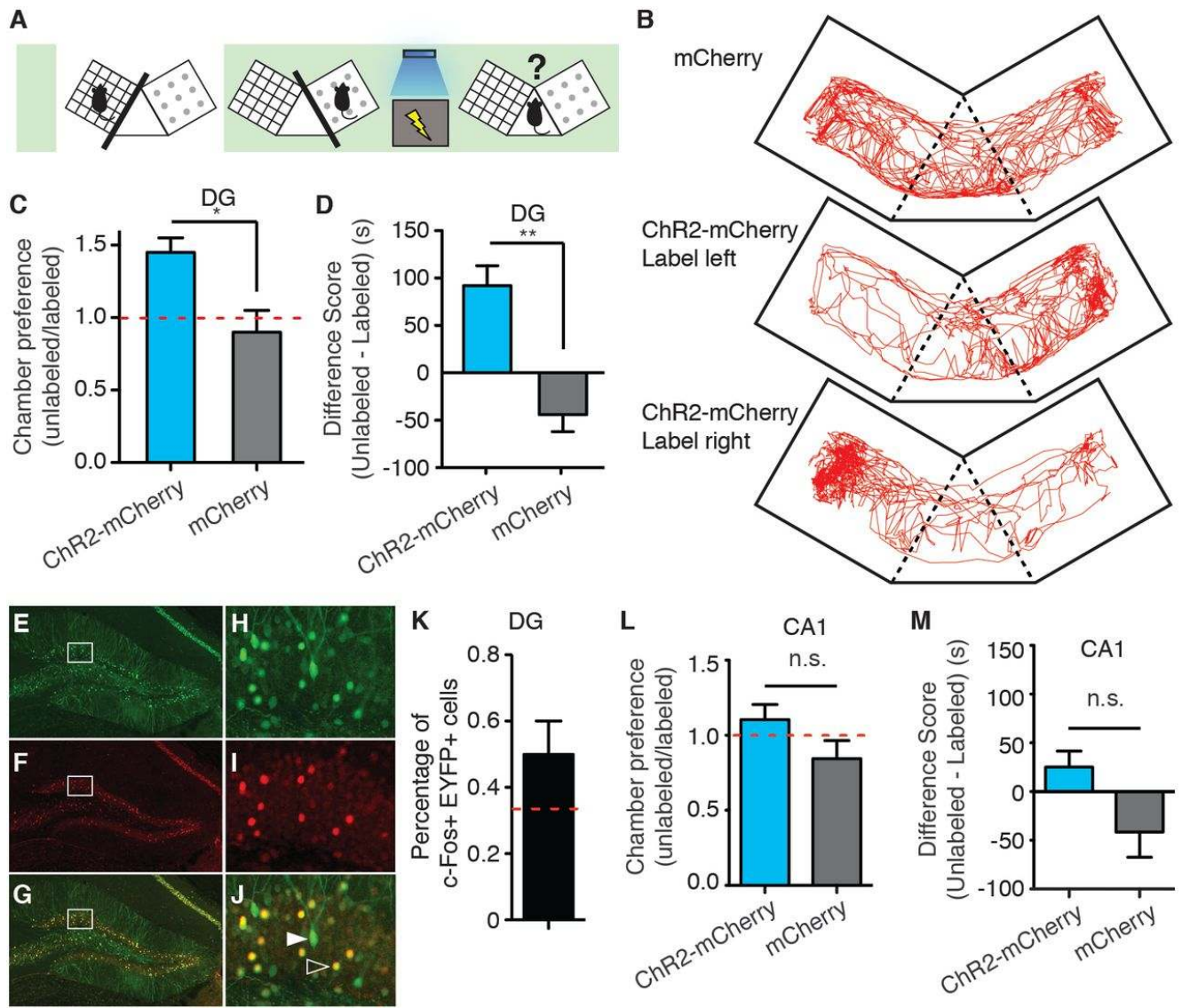
Figure 1

Figure 2

Figure 3

Figure 4