Peer Review File

Manuscript Title: Prefrontal feature representations drive memory recall

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

The current manuscript built upon the senior author's previous study (Rajasethupathy et al. 2015, Nature, Ref-21 in the manuscript), which identified the monosynaptic connection from anterior cingulate cortex (AC) to the CA3 and CA1 (CA) regions of the hippocampus and showed that this top-down AC-CA connection can drive fear memory retrieval. Using a similar approach, the current study added to those findings by showing correlative evidence that while CA1 neurons preferentially encode conjunctive representation of the context, AC neurons encode individual features of the context. The authors attempted to provide evidence to support their proposed model, in which feature representations encoded in AC drives contextual memory recall by activating the conjunctive representations in CA. In this reviewer's opinion, critical data to firmly establish the mechanistic model proposed are still missing. As such, the potential impact of the current study is significantly diminished.

Major Points

The main issues that need attention fall into two main areas: greater justification for the approach and greater supporting data.

1. The experimental (cognitive) paradigm is central to the conclusions presented in this article, and it is this reviewer's opinion that the manuscript requires greater explanation and justification of the cognitive task. This is especially true with respect to Figure 4, where the authors claim slight deviations from the initial task yet implement a significantly more salient aversive stimulus. The current presentation of the experimental paradigm left the following questions:

a. How translational is this type of learning to traditional hippocampal dependent tasks?

b. What visual stimuli are presented during training? (describe)

c. Are visual and tactile stimuli present during all probe conditions? If so, the design seems to be testing the alternating absence of olfactory and auditory stimuli. To this point, all stimuli are presented in conjunctive trials with at least 2 other stimuli. If you are only looking at compound feature representations can you really call them feature neurons?

This reviewer also found that with the current presentation of the data it was difficult to follow whether data presented were from training or retrieval. Greater delineation between the two data types, as well as the difference between context/feature-responsive and context/feature-selective neurons would greatly improve reader comprehension.

2. This reviewer believes that there is a need for additional rationale for the investigation of both brain regions presented in this study. Of note, the region here identified as the dorsal CA1 was previously characterized by these authors as dorsal CA2/CA3 [Ref-21]. While the authors present data from inhibition experiments that support the selection of the AC, the retrograde tracing experiment doesn't address the possibility that neocortical regions other than the AC contribute to

feature encoding in CA1 (as previously suggested by others [for example, Ref-37]). This point is echoed throughout the manuscript: there is a consistent need for greater reconciliation between the present study and previous literature. This is particularly true as the novelty of the study is questionable. It is well-established that the hippocampus is capable of pattern completion, and therefore the region has been shown to be capable of conjunctive sensory representation.

3. A major issue that needs to be addressed is the heavy reliance on correlational analysis. The vast majority of the data presented in each figure is from "representative" mice, trials, and cells. Considering the very small sample size of the study, it is necessary to report and show the mean/median and standard error of the reported data. Further, there is no clear justification for the abundance of analytical approaches relying heavily on categorical classification of relative values (eg. fractions of time active, "net relative response of a feature selective ensemble to other features"). More standard quantifications, such as regional reductions dF/F during inhibition, should be plainly presented and applicability of alternative approaches should be supported with literature (eg. t-tests of correlation coefficient values, population trajectories, support vector machine analysis).

4. Strong, non-correlational evidence establishing the functional connection between AC and CA is paramount to the conclusions in this study and is presently lacking. Evidence showing that feature-selective, but not other, AC neurons specifically activate context CA neurons and drive context-specific retrieval would be key to substantiating the proposed model.

5. There is a further need to address the role that saliency plays in the present study. The AC-CA connection's influence on retrieval depends on saliency, as shown in Fig. 4 and also in the previous study [Ref-21]. Contextual encoding does not depend on saliency, which is needed only for evaluation of the quality of contextual memory.

a. Is the feature encoding in AC dependent on saliency? Note that the neutral context was excluded in the first set of experiments (Figs. 1-3), and aversive context was compared to neutral context (as control) in the last set of experiments (Fig. 4). This inconsistency drew into question whether results from one set of experiments can be used to interpret those from another set of experiments.

b. If AC encoding is saliency dependent, an alternative interpretation of the findings would be that the AC-CA connection facilitates memory retrieval and AC may not be the (key) site for feature encoding.

Other Specific Points

Fig. 1

Fig. 1g – The 10% context selective neurons recorded in training is not consistent with what's stated in the text: "During training, we observed a substantial fraction (~20%) of neurons that responded selectively to the reward or aversive context (Fig. 1d,e,g)."

Fig. 1i – As stated earlier, can compound features be used to define feature neurons? How about single features?

Extended Data Fig. 4h - mislabeled

In addition to CA1, how about other hippocampal regions, like CA3 and DG? DG performs pattern separation and CA3 is for pattern completion. The DG receives dense inputs from lateral entorhinal cortex, "where both conjunctive and feature-like responses have been observed". It is reasonable that DG could be one region for feature representation.

Fig. 2

The figure caption, "AC, not LEC, provides feature inputs to CA1 during memory recall", is not substantiated by the data presented in the figure. All it showed was that inhibition of AC suppressed

the activity of CA1 context neurons and impaired CA retrieval, which has been shown previously [Ref-21]. In particular, there are conjunctive neurons in AC (Fig. 3) and the data presented don't provide any evidence that AC provides feature inputs to CA1.
2e-h – Although LEC inhibition doesn't preferentially inhibit context responsive neurons, the involvement of LEC in feature encoding cannot be ruled out.
2h – "n=3 mice, 6 sessions for LEC, data are mean with each mouse as individual data point". There are 6 points for LEC. Does one point represent one mouse, or one session? Same for AC.
2f-g – Please include statistics for the CDF plots.
2i – Please provide the lick rate during memory retrieval, not just discrimination index.
The PI's previous paper [Ref-21] showed AC also projects to CA2/3. How does CA2/3 respond to AC inhibition? Is the impaired retrieval due to the direct projection AC-CA1 or a compound effect?
Is AC more special than other cortical areas? LEC also controls context responsive neurons in CA1. How about MEC? If inhibition of LEC/MEC similarly impair memory retrieval (very likely), could you also conclude that LEC/MEC provides feature inputs to CA1?

Fig. 3
3c – Quantifications of context-selective neurons in training, feature-selective and conjunctive neurons in retrieval should be shown.
3d – The Venn diagram is a very nice presentation, but it is from only one mouse. Results from all mice should be shown. The Venn diagram presentation should be used for data from CA1 as well.
3f-g – Again, data from only one mouse is presented, and only the aversive context. Please label the x axis groups in 3f. What are the variabilities between animals? Are there any differences between the aversive and reward context? This is an excellent opportunity for repeated-measures ANOVA, or if limited by sample size, the different probes (AOT, AT, OT) could be averaged.
3g – Based on the data from the one representative animal, it seems that CA1 context neurons are generally more active than AC neurons during retrieval but also show a certain degree of feature selectivity. If the data are representative for all animals, doesn't it suggest that features are imbedded in the conjunctive representations in CA1?

3h – Same as 2h. It looks like one point represents one session. I do not believe it is appropriate to average mice into sessions and use those sessions as input for the ANOVA. How do the results look when the data are grouped by animal?
Direct evidence is needed to show that the feature encoding in AC is responsible for context retrieval in CA1.

Fig. 4
This figure does not specifically show the interaction between "the feature codes" in AC and CA1 representation. The AC context neurons are not classified by features here.
What is the rationale to focus on aversive and neutral contexts, not reward? Please point out that the aversive stimuli are changed from air puffs to electric shocks and explain why.
4d – "During training, we observed that the recruitment of CA1 context ensembles significantly preceded the recruitment of AC feature ensembles." "During retrieval, however, AC feature ensembles preceded CA1 ensembles." Are the reported statistics for all mice or only for the representative mouse? It is well established that memory is initially encoded in the hippocampus and gradually consolidated in neocortex. Can this simply explain the observed phenomenon, due to

the time window for training and retrieval?
4f-h – What specific criteria was used to select “long range neurons”? What percentage of “long range neurons” come from each group (Neutral-Neutral, Aversive-Aversive)? Note the correlation is much higher in aversive context. The “long range neurons” comprising largely aversive ensembles entirely explains the data presented in g and h.
4i-j – Are the middle panels the “quantification of net dF/F change (ITI to context onset)” as stated? This quantification has no relevance to the conclusions drawn from this data.
The data presented here show that the shock neurons gain context-selectivity during training in CA1 and AC. However, there is no direct evidence supporting the bi-directional interaction between CA1 and AC. Does inhibition of CA1 neurons impair context -selectivity in AC shock neurons during training? Again, no feature encoding is measured here.

Referee #2 (Remarks to the Author):

In this manuscript, Yadav et al. describe the results of an exciting new study that identifies the locus of contextual feature representation in the anterior cingulate cortex that recruits hippocampal contextual ensembles after learning. The authors use a multidisciplinary approach, combining cutting edge longitudinal 2-photon imaging in both the hippocampus and the anterior cingulate cortex (AC) with optogenetic manipulation to understand how the encoding of complex, conjunctive contextual memories are driven by representations of their constituent features. The authors show that the AC, but not the entorhinal cortex, can drive contextual ensembles in the CA1 to modulate memory retrieval. Importantly, they determine that while the hippocampus codes for a complex, conjunctive representation of a context, its component features are encoded by the AC, thus identifying a parallel but qualitatively distinct memory code from that in the CA1. While the manuscript is logically structured, well written and experiments have been well designed; it may benefit from addressing the following concerns:
Main points:
1. Fig. 1a shows (and the authors have noted) that the mice cannot distinguish between the aversive and neutral context. Given this lack of discrimination in the earlier results, it is important to determine whether the mice could distinguish between the aversive and neutral context in the modified behavioral paradigm used for dual region imaging (Fig.4). This will also strengthen the authors observations of the importance of salience in developing hippocampal-neocortical interactions.
2. In Fig. 2b-h, the authors show that inhibition of AC, but not LEC neurons during retrieval, leads to an inhibition of context neurons in the CA1. However, since this AC inhibition was somatic and not localized to AC afferents in the CA1, it is not clear whether the specific inhibition of CA1 context neurons was driven by direct projections from the AC to CA1 or through a broader network. One way to answer this question would be to determine whether CA1 projecting AC, but not LEC neurons, preferentially synapse onto context selective CA1 neurons. This is especially interesting given the author’s finding that the AC and LEC synapse onto non-overlapping subsets of neurons in the CA1 (Extended data Fig. 7).
3. In Fig. 4d, the authors describe a very interesting and exciting temporal coupling between the CA1 and AC context ensembles that shifts from training to retrieval. Does this coupling detach or lose

power in retrieval trials where the mouse fails to recognize the context? This analysis could help to establish a causal link between these AC-CA1 interactions and contextual memory recall.
Other points:
1. The authors should clearly explain the reason for shifting their behavioral paradigm from a reward versus aversive test in Figs. 1-3 to an aversive versus neutral paradigm in Fig. 4.
2. What is the threshold value used to identify the 'highly correlated' neurons across the AC-CA1 in Fig. 4f-h?
3. Fig. 4d demonstrates a temporal coupling between the AC and CA1 that changes in hierarchy across the learning period. Is this shift in hierarchy tied to the emergence of feature and context selective neurons in the AC and CA1 respectively across the learning duration? In other words, do context selective neurons emerge in the CA1 earlier in the learning process and subsequently play a role in instructing the development of feature selective neurons in the AC?
4. If the above hypothesis holds true, can you block the development of feature selective neurons in the AC by blocking AC-CA1 interactions during training?
In summary: this manuscript reports exciting findings that address an important question and do so with state-of-the-art imaginative experiments. I only have a few concerns and suggestions that the authors should address before publication

Author Rebuttals to Initial Comments:

Response to Reviewer Comments:

We thank the reviewers for taking the time and effort to provide detailed, constructive, feedback. We were encouraged by the shared enthusiasm for the overall approach and findings, i.e, "this manuscript reports exciting findings that address an important question and do so with state-of-the-art imaginative experiments." However, both reviewers requested additional justification of approach and inclusion of data to support our proposed model. We agree with all the points raised by the reviewers and we have now performed 5 additional experiments to address them, including experiments related to the hippocampal dependency of the task (ED Fig. 1c), causal role of LEC in behavioral recall (Fig. 2h,i), whether the emergence of AC representations requires CA1 (ED Fig. 8m-o), anatomical tracing of ACC/LEC projections onto context neurons in CA1 (ED Fig. 7d,e), and finally, and perhaps most striking, a full and new dataset demonstrating *near complete* deficit in feature based recall when inhibiting preferentially the feature-coding neurons (but not another similarly sized set of neurons in AC) (Fig. 3j,k; ED Fig. 9f). We have also clarified aspects of the behavioral design and statistical analyses, and to the extent possible, have addressed other technical and minor concerns through either inclusion of new data, new analyses of existing data, and/or revisions of the text to provide more precision and clarity. We provide a detailed point-by-point response to reviewers' comments below.

We believe the new manuscript is greatly strengthened as a result of this revision, and we are very thankful for the thoughtful comments and support from the Reviewers and the Editor.

As a side note: We are aware that a recent non-peer reviewed paper was uploaded on bioarxiv claiming they were not able to find the AC to CA1/CA3 projection: https://doi.org/10.1101/2022.01.25.477805. We note here that there are several technical flaws in their study, most notably: 1) Their retrograde tracing (Fig. 2) is targeted to DG (many yellow starter cells) not CA3/CA1 (<5 yellow cells in CA1 & 0 in CA3). And 2) their anterograde tracing (Fig. 1) is weak as they are not able to identify positive controls such as strong projections to MD. To reduce technical variance, we have recommended to anyone that has contacted us to simply use cre in both directions (AAV1-cre in anterograde or AAVretro-cre in retrograde) and they have all been able to replicate this pathway robustly. In this manuscript, we provided examples of our tracings in Fig. 2, ED Fig 5 (retrograde, no cre used) and ED Fig 7 (anterograde, using cre), which we further confirmed are robust to type of virus (AAV, retroAAV, CAV, Rabies), serotype, or promoter (Ef1a, CAG, hSyn, CaMKII) used. Another lab has also described monosynaptic PFC to hippocampus projections: https://www.biorxiv.org/content/10.1101/2021.03.01.433441v1, which were also missed by the Craig Lab.

Reviewer #1

Major Points:

1. The experimental (cognitive) paradigm is central to the conclusions presented in this article, and it is this reviewer's opinion that the manuscript requires greater explanation and justification of the cognitive task. This is especially true with respect to Figure 4, where the authors claim slight deviations from the initial task yet implement a significantly more salient aversive stimulus. The current presentation of the experimental paradigm left the following questions:

a. How translational is this type of learning to traditional hippocampal dependent tasks?

Traditional hippocampal dependent tasks have a strong spatial or contextual component. Since it has been well documented that single cue-associations are not hippocampal dependent whereas multi-modal context associations are hippocampal dependent, we deliberately chose a multi-modal contextual association task to ensure hippocampal dependency and to relate with the existing literature. However, as the reviewer notes, it is important to directly demonstrate the hippocampal dependency of this task. Thus, we now have performed bilateral inhibition of CA1 during the training phase of this task. We demonstrate that CA1 inhibited mice have strong deficits in learning the reward and aversive context associations compared to the non-inhibited control cohort (ED Fig. 1C). We thank the reviewer for encouraging us to clarify and include this.

b. What visual stimuli are presented during training? (describe)

We have revised the task description to now include more details as recommended by the reviewer. The visual stimuli for each of the three contexts consist of rooms with differing colors, shapes, and patterns which together provide differences in visual contrast on the virtual floor and walls as they travel through the rooms. The reward room has white rectangles on a pink floor and white circles on pink walls; the neutral room has white rectangles on blue floors and tiled rectangles on blue walls, and the aversive room has white rectangles on black floors and white octagons on black walls (Supp Video). These stimuli were chosen either based on previous literature or our own prior empirical experience on what the mice are able to discriminate. These details are now included in Methods P. 4.

c. Are visual and tactile stimuli present during all probe conditions? If so, the design seems to be testing the alternating absence of olfactory and auditory stimuli. To this point, all stimuli are presented in conjunctive trials with at least 2 other stimuli. If you are only looking at compound feature representations can you really call them feature neurons?

Visual and tactile stimuli are not presented on all probe conditions, only the feature type being tested is presented (so AT would include only auditory and texture; O would include only olfactory etc). We apologize for the lack of clarity and have revised the task description (in Main text P. 3 and Methods) to make this more explicit.

Regarding single vs. compound features, we initially designed the task to be able to probe all possible single, two-cue, and three-cue feature types, but this amounted to 18 trial types, which were too many to test and appropriately power. We thus reduced our feature set to 4 types (AT, OT, AOT, AVOT) which allowed us to test many more trials of each trial type and increase power of these studies. We chose to focus on feature types consisting of 2, 3, and 4 cues because these resulted in less variable behavioral and neural responses across mice, though single cue responses were also observed. Interestingly, in almost every case, we found that ACC neurons responding to compound cues (i.e., AOT) were not summations of responses to single cues (A, O, and T) or dual cues (AT and OT, Fig. 3e), in fact completely non-overlapping sets of neurons represented each single and compound cue type; Even behaviorally, mice that performed recall using compound cues (i.e., OT) were not necessarily able to do so using cognate single cues (O and T, Fig. 1b) thus supporting the idea that each stimulus set (single or compound) is represented by the animal as a separate feature type.

nature portfolio

Finally, in response to the reviewer's concern "*the authors claim slight deviations from the initial task yet implement a significantly more salient aversive stimulus"* we would like to clarify that there were two main reasons to move to this new behavior for Fig 4: Firstly, given the technical challenges of dual region imaging and tracking the same neurons in both regions across days, we wanted to reduce the task to one day of training, and one day of retrieval. By extending the time in context on the training day, and switching reinforcement from airpuff to electric shock, we observed mice reliably learn this task within a day which allowed us to track the relevant neurons in both regions reliably across the two days period (rather than the 6 day period in the previous version of the task). Secondly, we wanted to understand the role of saliency and we could have chosen to study reward vs neutral (instead of aversive vs neutral) but we chose aversive vs neutral because it was easier to train the mice to perform context discrimination within one day. We have added these points to the Main text (p. 8) and Methods to increase clarity and transparency for the behaviors chosen and implemented.

This reviewer also found that with the current presentation of the data it was difficult to follow whether data presented were from training or retrieval. Greater delineation between the two data types, as well as the difference between context/feature-responsive and context/feature-selective neurons would greatly improve reader comprehension.

As suggested by the reviewer, we have systematically gone through the manuscript (in both main text and figures) and updated the manuscript to clearly delineate the training and retrieval phases of the task. We have also added more clarity to the definitions for "feature/context responsive neurons" vs. "feature/context selective neurons". Briefly, feature responsive neurons are neurons that reliably respond to that feature (but may also respond to other features of the same context). By comparison, feature-selective neurons exhibit significant responses only to that feature but not other features of the same context. The exact statistical definitions are provided in the Methods section.

2. This reviewer believes that there is a need for additional rationale for the investigation of both brain regions presented in this study. Of note, the region here identified as the dorsal CA1 was previously characterized by these authors as dorsal CA2/CA3 [Ref-21]. While the authors present data from inhibition experiments that support the selection of the AC, the retrograde tracing experiment doesn't address the possibility that neocortical regions other than the AC contribute to feature encoding in CA1 (as previously suggested by others [for example, Ref-37]).

As noted by the reviewer, direct projections from AC to both CA1 and CA3 have been documented. We chose to focus on CA1 in this study because it has been extensively shown to be required for almost every type of contextual memory retrieval process, but there is limited understanding of whether it represents only the global context, or also the constituent features. Upon finding chance level representations of feature-selective responses in CA1 (Fig. 1g), we aimed to understand what brain regions route feature information to CA1 to enable memory recall, and focused on two regions, AC and LEC, that have monosynaptic projections to CA1.

Indeed Ref 37 and related studies encouraged us to consider lateral entorhinal cortex and other cortical regions (rather than intra-hippocampal regions such as DG which route

nature portfolio

through CA3, where the recurrent architecture is likely to create conjunctive responses rather than maintain feature responses). We now include this discussion in Main text P. 4.

While we don't refute that AC may not be the only feature-coding region and that other neocortical regions may contribute, we were surprised that when comparing two neocortical regions, only AC inhibition (not LEC inhibition) led to deficits in recruitment of context neurons in CA1 during recall (Fig. 2d-g). We have now also added new data, encouraged by this reviewer's comment, to more directly assess contributions of LEC to behavioral recall. We performed bilateral inhibition of LEC during the retrieval phase of this task and found no significant deficits in feature-based recall (Fig. 2h) in contrast to the significant deficits observed with bilateral AC inhibition (Fig. 2i). One interpretation of these findings, in the context of the broader literature, could be that LEC has more prominent functions in storage whereas ACC functions more prominently in retrieval, a model that is further supported by recent and emerging studies[1-3]. We thank the reviewer for encouraging us to perform these experiments, which indeed strengthens the proposed model. Whether the feature coding in AC is shaped by other neocortical regions could be an interesting view for a follow up study.

3. A major issue that needs to be addressed is the heavy reliance on correlational analysis. The vast majority of the data presented in each figure is from "representative" mice, trials, and cells. Considering the very small sample size of the study, it is necessary to report and show the mean/median and standard error of the reported data. Further, there is no clear justification for the abundance of analytical approaches relying heavily on categorical classification of relative values (eg. fractions of time active, "net relative response of a feature selective ensemble to other features"). More standard quantifications, such as regional reductions dF/F during inhibition, should be plainly presented and applicability of alternative approaches should be supported with literature (eg. t-tests of correlation coefficient values, population trajectories, support vector machine analysis).

We thank the reviewer for raising these points.

Regarding representative vs summary figures, wherever possible we tried to show examples of raw data from representative mice to build confidence in the results and to provide an intuition to interpret the summary statistics. However, as the reviewer notes, we found some figure panels where the summary data and statistics were not provided and we have now combed through all of the figures and made certain to include mean/median and standard deviation/standard error of mean where applicable, along with the associated statistical tests performed. For instance, all of Fig. 4 (together with corresponding ED Fig. 10) now has statistics shown for each individual mouse and also aggregate summary statistics for all mice.

Regarding quantifications, we have switched to using standard quantifications wherever possible, but in some cases where categorical representations are required (detection of GCaMP onsets, correlated pairs) or normalization is needed (to compare behavioral performance across cohorts) we have added citations to their prior use in the literature, either in Main text or Methods. For Fig. 2e-g, we would like to clarify that for identifying inactive neurons in CA1, we used changes in the binarized activity of the dF/F (which is commonly used to detect GCaMP <u>onsets</u>[4-6], ED Fig. 2,8) rather than dF/F <u>magnitude</u> simply because the GCaMP activity is very variable in the *light off* trials which can heavily bias the extent of inhibition if calculating as a ratio of *light on dF/F* to *light off dF/F*. Thus, quantifying the change in <u>timing</u> rather than <u>magnitude</u> of responses was a more accurate and

nature portfolio

conservative approach to quantifying extent of inhibition. Indeed using this approach we were able to reduce the variability of *light off* trials across trials and mice (ED Fig. 6e,f). However, we have now also included a separate analysis quantifying magnitude of dF/F reductions during *light on* trials and observe largely similar results (ED Fig. 6k). These explanations and corresponding citations have been added to the Methods section. We hope these modifications have strengthened the statistical interpretations.

4. Strong, non-correlational evidence establishing the functional connection between AC and CA is paramount to the conclusions in this study and is presently lacking. Evidence showing that feature-selective, but not other, AC neurons specifically activate context CA neurons and drive context-specific retrieval would be key to substantiating the proposed model.

We agree that this experiment provides important support for the proposed model. We therefore performed an experiment to preferentially inhibit feature-coding neurons in AC during memory recall. Using FosTRAP2 mice[7,8], which we characterized using multiple protocols and optimized, we find impressively that preferential silencing of the feature neurons in AC (but not a similarly sized set of other neurons) is sufficient to produce almost complete deficit in feature based recall.

Specifically, we injected the cre-dependent inhibitory opsin SIO-stGtACR into FosTRAP2 mice and provided tamoxifen on retrieval day 1, during feature presentations only, to preferentially express st-GtACR in feature-coding neurons of AC (~20% of all neurons, Fig. 3j, ED Fig. 9f). We found that optical silencing of feature-neurons resulted in almost complete deficit in feature-based recall (Fig. 3k left panel), which was rescued by providing no optical inhibition the following day (Fig. 3k left panel; $F(1,25)=64.39$, $p<0.0001$; Two-way ANOVA between R6 optical inhibition vs R7 no optical inhibition). In contrast, expression of the inhibitory opsin in a random population of neurons of a similar size (also ~20%, which we obtained by applying tamoxifen on the last day of habituation prior to training on task, Fig. 3j, ED Fig. 9f) and silencing during retrieval resulted in no significant deficit in feature-based recall (Fig 3k right panel; $F(1,20)=0.14$; $p>0.05$; Two-way ANOVA between R6 inhibited vs R7 non-inhibited). These results demonstrate a strong causal role for feature-coding neurons in driving memory recall.

We additionally add support that AC has high specificity in targeting CA1 context neurons during recall by performing AC/LEC anterograde tracing and overlaying with cFos immunostaining subsequent to behavioral recall (which preferentially labels context neurons in CA1). We found significantly greater overlap of CA1 cFos+ context neurons with ACC-projecting (compared to LEC projecting) partners in CA1, supporting a direct anatomical interaction between AC and CA1 mediating recall (ED Fig. 7d,e). These additions have greatly strengthened the proposed model and we thank the reviewer for these suggestions.

5. There is a further need to address the role that saliency plays in the present study. The AC-CA connection's influence on retrieval depends on saliency, as shown in Fig. 4 and also in the previous study [Ref-21]. Contextual encoding does not depend on saliency, which is needed only for evaluation of the quality of contextual memory.

a. Is the feature encoding in AC dependent on saliency? Note that the neutral context was excluded in the first set of experiments (Figs. 1-3), and aversive context was compared to neutral context (as control) in the

nature portfolio

last set of experiments (Fig. 4). This inconsistency drew into question whether results from one set of experiments can be used to interpret those from another set of experiments.
b. If AC encoding is saliency dependent, an alternative interpretation of the findings would be that the AC-CA connection facilitates memory retrieval and AC may not be the (key) site for feature encoding.

We agree that our interpretation and discussion of the AC dependence on saliency was not clear and we have attempted to explain this better. Briefly, in the first set of experiments (Figs. 1-3), we excluded the neutral context in order to maximize the number of trials of reward and aversive feature presentations (otherwise we noticed we were not able to sufficiently power the study if including all neutral feature conditions as well). These analyses allowed us to conclude that feature representations exist in AC regardless of the valence of the associated context.

Next in Fig 4, as suggested by the reviewer, we also wanted to delineate the effect of saliency on memory recall. Thus, we implemented a version of the task designed to compare the aversive vs neutral context and found prominent AC feature representations, and AC-CA1 synchrony, in both the salient and non-salient context (Fig. 4d, correlations in neutral context are significantly above chance; where chance Pearson=0.1±0.07). Thus neutral contexts are indeed represented in AC and in AC-CA1 interactions, however, we found that saliency further enhances the bidirectional communication (AC-CA correlations are even higher in the aversive context). We have added this to the Main text P.8 for clarification.

Other Specific Points

Fig. 1
Fig. 1g – The 10% context selective neurons recorded in training is not consistent with what's stated in the text: "During training, we observed a substantial fraction (~20%) of neurons that responded selectively to the reward or aversive context (Fig. 1d,e,g)."
Sorry this is a typo and we have corrected it.

Fig. 1i – As stated earlier, can compound features be used to define feature neurons? How about single features? As discussed above, we find that each cue combination (single, dual, triple) has separate (not overlapping or additive) neural representation and behavioral significance, thus we consider each a separate feature type.
Extended Data Fig. 4h - mislabeled Corrected, thank you.

Fig. 2
The figure caption, "AC, not LEC, provides feature inputs to CA1 during memory recall", is not substantiated by the data presented in the figure. All it showed was that inhibition of AC suppressed the activity of CA1 context neurons and impaired CA retrieval, which has been shown previously [Ref-21]. 2e-h – Although LEC inhibition doesn't preferentially inhibit context responsive neurons, the involvement of LEC in feature encoding cannot be ruled out.

We agree with the Reviewer's point here. We have now included additional experiments involving LEC inhibition, which did not suppress behavioral feature-based recall (Fig. 2h,i), and we have changed the title of this figure caption to reflect the new data and in line with the Reviewer's point to state: "AC, but not LEC, is required for feature-based recall". Of

nature portfolio

note, and encouraged by the reviewer, we have added additional experiments to Fig. 3 which also speaks to this point and more convincingly demonstrates that, within AC, it is specifically the feature-coding neurons in AC that are required for recall (Fig. 3j,k). Also to clarify, Ref 21 did not include any experiments involving simultaneous inhibition of AC while imaging in CA, which we hope within the context of feature-presentations here, has clarified the functional roles of AC in retrieval.

2h – "n=3 mice, 6 sessions for LEC, data are mean with each mouse as individual data point". There are 6 points for LEC. Does one point represent one mouse, or one session? Same for AC.

Each data point represents a session (n=6), but the statistics have been performed using numbers of mice (n=3), we have clarified this in the legend and added ED Fig. 6j.

2f-g – Please include statistics for the CDF plots.

We have now included statistics in the figure and their description in the legend.

2i – Please provide the lick rate during memory retrieval, not just discrimination index.
During retrieval (as opposed to training) there is greater variability in raw lick rates across mice because different mice have preferences to different features (as discussed in Main text P.3 and shown in Fig. 1b). Thus, to combine data across mice we used a normalized lick rate discrimination index (described in Methods), which is used widely in the literature as a metric of learning and performance[9-11].

Is AC more special than other cortical areas? LEC also controls context responsive neurons in CA1. How about MEC? If inhibition of LEC/MEC similarly impair memory retrieval (very likely), could you also conclude that LEC/MEC provides feature inputs to CA1?

We have added LEC inhibition data to address this point (Fig. 2h). We too were surprised that LEC inhibition did not affect recruitment of context neurons in CA1 (Fig. 2e-g), and now with new data we find LEC inhibition does not significantly affect feature-based recall (Fig. 2h). We thus suggest a parsimonious model that perhaps LEC/MEC inputs are more important for storage whereas the AC inputs are more important for retrieval (Fig. 4h, P.9), and recent studies support this model[1-3].

Fig. 3
3c – Quantifications of context-selective neurons in training, feature-selective and conjunctive neurons in retrieval should be shown.
We have now provided this in Fig. 3c.

3d – The Venn diagram is a very nice presentation, but it is from only one mouse. Results from all mice should be shown. The Venn diagram presentation should be used for data from CA1 as well.
When aggregating data for the venn diagram, we noticed variability across mice that we thought might be best captured in violin plots, which we have now included. Thank you.

3f-g – Again, data from only one mouse is presented, and only the aversive context. Please label the x axis groups in 3f. What are the variabilities between animals? Are there any differences between the aversive and

nature portfolio

reward context? This is an excellent opportunity for repeated-measures ANOVA, or if limited by sample size, the different probes (AOT, AT, OT) could be averaged.

We apologize for the confusion we created on these panels. These are cumulative data from n=3 mice; the heatmap only has ~300 neurons because we are only showing the context-responsive neurons, ~10% of all 3000 recorded neurons). We have now corrected the legend to accurately describe what is shown.

Only the aversive context data are shown due to space constraints but the reward context data are very similar and are now shown in ED Fig. 9b.

3g – Based on the data from the one representative animal, it seems that CA1 context neurons are generally more active than AC neurons during retrieval but also show a certain degree of feature selectivity. If the data are representative for all animals, doesn't it suggest that features are imbedded in the conjunctive representations in CA1?

We thank the reviewer for raising this. We agree that the visualization of the heatmaps (which, as mentioned above, is from 3 mice) suggests feature coding embedded within context coding in CA1, which we realized is because the heatmap data are sorted across rows. Such sorting would give a visual impression of tiling even in a random dataset. We thus modified the figure in two ways to correct this misleading aspect of the display: (a), we now present the heatmaps without sorting, and (b) we add a black/white lane to the right, showing every neuron that was classified as a feature-selective neuron as a white bar (Fig. 3g). For CA1, the number of neurons classified as feature selective is at chance levels (Fig. 3g, 1g). This shows that the distinctions between context encoding in CA1 vs feature encoding in AC are captured not only at the single neuron but also at the population level.

We also would like to note that the distinction between context and feature coding is not due to the imposition of a categorical analysis on the data: it is also apparent in the coding-space analysis of Fig. 3h, where there is no thresholding.

3h – Same as 2h. It looks like one point represents one session. I do not believe it is appropriate to average mice into sessions and use those sessions as input for the ANOVA. How do the results look when the data are grouped by animal?

We have included these same data grouped by animal (ED Fig. 9c and ED Fig. 6j), and performed statistics by animal rather than session, and demonstrate similar statistical significance described in legends.

Fig. 4

This figure does not specifically show the interaction between "the feature codes" in AC and CA1 representation. The AC context neurons are not classified by features here.

We classified the AC neurons into feature categories and then pooled all of these into a single "feature coding ensemble" that we use in all analyses throughout this figure. We have made this more explicit in the figure labels and legends.

What is the rationale to focus on aversive and neutral contexts, not reward? Please point out that the aversive stimuli are changed from air puffs to electric shocks and explain why.

nature portfolio

As described above, a major goal with the mesoscope experiments was to reduce the training to one day, which was the reason we switched the reinforcement from airpuff to shock. Another major goal was to understand the role of saliency- as the reviewer rightly notes, we could have compared reward vs neutral (instead of aversive vs neutral) but we chose aversive because it was easier to train the mice to perform context discrimination in one day. We have now added these clarifications to the manuscript P. 8.

4d – "During training, we observed that the recruitment of CA1 context ensembles significantly preceded the recruitment of AC feature ensembles." "During retrieval, however, AC feature ensembles preceded CA1 ensembles." Are the reported statistics for all mice or only for the representative mouse? It is well established that memory is initially encoded in the hippocampus and gradually consolidated in neocortex. Can this simply explain the observed phenomenon, due to the time window for training and retrieval?

We have shown a representative mouse in Fig. 4, but this result holds for each mouse tested (ED Fig. 10), and we now provide summary statistics for all mice in Fig. 4c. Indeed it has been widely believed that memory is gradually consolidated in neocortex, but this is thought to occur over weeks in rodents, whereas we observe this bi-directional change between AC and CA1 interactions within a day.

4f-h – What specific criteria was used to select "long range neurons"? What percentage of "long range neurons" come from each group (Neutral-Neutral, Aversive-Aversive)? Note the correlation is much higher in aversive context. The "long range neurons" comprising largely aversive ensembles entirely explains the data presented in g and h.

In the method section, we note "Highly-connected long-range neurons in both regions were defined as those neurons for which the number of correlated partners exceeded the average of the neurons in the same volume by 1 standard deviation". We have now included this in the main text for clarity, as this confusion was also raised by reviewer 2. Indeed the reviewer's conclusion *"long range neurons comprise largely of aversive ensembles"* is the exact same conclusion we are making (taking an unbiased approach by defining long range neurons first, independent of response selectivity), but the reviewer has stated it more clearly and we have now adopted the same phrasing on P. 8/9.

4i-j – Are the middle panels the "quantification of net dF/F change (ITI to context onset)" as stated? This quantification has no relevance to the conclusions drawn from this data.

The legend and x-axis label are incorrect here. The x-axis was given in frames, and now we have changed it to seconds. Thus, the legend now reads "quantification of dF/F response (dF/F during first 10s in context normalized to ITI)".

The data presented here show that the shock neurons gain context-selectivity during training in CA1 and AC. However, there is no direct evidence supporting the bi-directional interaction between CA1 and AC. Does inhibition of CA1 neurons impair context -selectivity in AC shock neurons during training?

We performed additional experiments that address this. We injected inhibitory hM4D(Gi) DREADD receptors into CA1 and provided i.p. CNO during task learning, while performing simultaneous imaging of AC. We found that silencing of CA1 led to a strong and

nature portfolio

significant loss of AC context decoding in hM4DI mice (ED Fig. 8m,n, 0.78 for hM4Di vs 0.56 for mCherry max decoding accuracy, P=0.008, Mann Whitney U Test on the distributions), compared to control mice also receiving CNO but lacking the inhibitory DREADD receptor.

Overall, we sincerely thank this reviewer for a detailed reading of the manuscript, and for the many helpful suggestions on experiments and text that have significantly strengthened and clarified the message and improved the quality of the manuscript.

Reviewer #2

Main points:

1. Fig. 1a shows (and the authors have noted) that the mice cannot distinguish between the aversive and neutral context. Given this lack of discrimination in the earlier results, it is important to determine whether the mice could distinguish between the aversive and neutral context in the modified behavioral paradigm used for dual region imaging (Fig.4). This will also strengthen the authors observations of the importance of salience in developing hippocampal-neocortical interactions.

We thank this reviewer for catching this important oversight. We have now included the behavior data of the mice that underwent dual region imaging in ED Fig. 10c, which show that mice do make this context distinction (mean lick rate =3.9Hz for neutral vs 2.3Hz for aversive context, P<0.05 paired t-test).

2. In Fig. 2b-h, the authors show that inhibition of AC, but not LEC neurons during retrieval, leads to an inhibition of context neurons in the CA1. However, since this AC inhibition was somatic and not localized to AC afferents in the CA1, it is not clear whether the specific inhibition of CA1 context neurons was driven by direct projections from the AC to CA1 or through a broader network. One way to answer this question would be to determine whether CA1 projecting AC, but not LEC neurons, preferentially synapse onto context selective CA1 neurons. This is especially interesting given the author's finding that the AC and LEC synapse onto non-overlapping subsets of neurons in the CA1 (Extended data Fig. 7).

We agree with this point and thank the reviewer for suggesting this experiment. To address this, we performed anterograde tracing from both AC and LEC to identify mono-synaptically connected neurons in CA1. We quantified the percentage overlap of either population with cFos+ neurons in CA1 that were identified by immunostaining 1.5h after the end of retrieval (which should preferentially label context neurons in CA1). We found significantly greater overlap of cFos+ neurons with AC-projecting (compared to LEC projecting) partners in CA1, supporting a direct anatomical interaction between AC and CA1 mediating recall (ED Fig. 7d,e, 39% overlap for AC vs 19% overlap for LEC projecting partners in CA1, n=4 mice each, P<0.01, Welch's t-test). We agree this is especially interesting in the context of an almost entirely non-overlapping population of AC vs LEC partners in CA1, and have added this in the text. We thank the reviewer for suggesting this experiment that provides important support to the proposed model.

3. In Fig. 4d, the authors describe a very interesting and exciting temporal coupling between the CA1 and AC context ensembles that shifts from training to retrieval. Does this coupling detach or lose power in retrieval trials where the mouse fails to recognize the context? This analysis could help to establish a causal link between these AC-CA1 interactions and contextual memory recall.

nature portfolio

This is an interesting idea. We unfortunately could not try this with the current dataset because in the Fig. 4 version of the task (explained further below & in Methods), there is only one day of training, with one long trial of each trial type, so we are not able to identify error trials. This is something we hope to consider in future studies.

Other points:

1. The authors should clearly explain the reason for shifting their behavioral paradigm from a reward versus aversive test in Figs. 1-3 to an aversive versus neutral paradigm in Fig. 4.

We apologize for the lack of clarity here and have now included detailed explanation in the Main text (P. 8). Briefly, there were two main reasons to move to this new behavior for Fig 4: Firstly, Given the technical challenges of dual region imaging and tracking the same neurons in both regions across days, we wanted to reduce the task to one day of training, and one day of retrieval. By extending the time in context on the training day, and switching reinforcement from airpuff to electric shock, we observed that mice reliably learn this task within a day which allowed us to track the relevant neurons in both regions reliably across the two days period (rather than the 6 day period in the previous version of the task). Secondly, we wanted to understand the role of saliency and we could have chosen to study reward vs neutral (instead of aversive vs neutral) but we chose aversive vs neutral because it was easier to train the mice to perform context discrimination within one day. We have added these points to the Main text (p. 8) to increase clarity and transparency for the behaviors chosen and implemented.

2. What is the threshold value used to identify 'highly correlated' neurons across the AC-CA1 in Fig. 4f-h?

In the method section, we note "Highly-connected long-range neurons in both regions were defined as those neurons for which the number of correlated partners exceeded the average of the neurons in the same volume by 1 standard deviation". This is performed by calculating pairwise correlations across AC and CA1 neurons. Pairs of neurons above a threshold cutoff of 0.3 were considered functionally connected (since electrophysiological studies have indicated a greater than 50% chance of in vivo functional connectedness when GCaMP correlations exceeds 0.3)[12,13]. To ensure that our results were robust to variations in this threshold, we repeated these analyses with Pearsons cutoff lower/bove this threshold as well and found consistent results. We have included this clarification in main text.

3. Fig. 4d demonstrates a temporal coupling between the AC and CA1 that changes in hierarchy across the learning period. Is this shift in hierarchy tied to the emergence of feature and context selective neurons in the AC and CA1 respectively across the learning duration? In other words, do context selective neurons emerge in the CA1 earlier in the learning process and subsequently play a role in instructing the development of feature selective neurons in the AC? 4. If the above hypothesis holds true, can you block the development of feature selective neurons in the AC by blocking AC-CA1 interactions during training?

We performed additional experiments to understand whether the hippocampal conjunctive representation supports AC feature representation or whether they emerge independently. We injected inhibitory hM4D(Gi) DREADD receptors into CA1 and provided i.p. CNO during task learning, while performing simultaneous imaging of AC. We found that silencing of CA1 led to a strong and significant loss of AC context decoding in hM4Di mice

nature portfolio

compared to control mice also receiving CNO but lacking the inhibitory DREADD receptor (ED Fig. 8m,n, 0.78 for hM4Di vs 0.56 for mCherry max decoding accuracy, P=0.008, Mann Whitney U Test on the distributions). It also led to a loss of AC feature representations (ED. Fig. 8o, 5.4 vs. 2.6 percent of feature coding neurons), where the effects trended in the expected direction but did not reach significance due to the small cohort. We unfortunately lost more mice than expected through the course of the experiment due to characterizations we performed to ensure efficacy of DREADD inhibition and due to poor GCaMP expression in some mice precluding imaging. Despite this, the deficit in context decoding and feature representations are striking, and we thank the reviewer for encouraging us to do this experiment. We intend to follow up on these results more thoroughly in future studies.

Overall, we sincerely appreciate this reviewer's considerate and thorough review, helpful suggestions, and for recognizing the challenges and novelty in the approach and findings.

References:

1. Lee JY, Jun H, Soma S, Nakazono T, Shiraiwa K, Dasgupta A, Nakagawa T, Xie JL, Chavez J, Romo R, Yungblut S, Hagihara M, Murata K, Igarashi KM. Dopamine facilitates associative memory encoding in the entorhinal cortex. Nature 598(7880):321-326 (2021).
2. Grienberger C, Magee JC. Entorhinal cortex directs learning-related changes in CA1 representations. bioRxiv 2021.12.10.472158.
3. Lopez-Rojas J, de Solis CA, Leroy F, Kandel ER, Siegelbaum SA. A direct lateral entorhinal cortex to hippocampal CA2 circuit conveys social information required for social memory. bioRxiv 2021.04.15.440048.
4. Dombeck,D. A., Harvey, C. D., Tian, L., Looger, L. L.&Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. Nature Neurosci. 13, 1433–1440 (2010).
5. Rajasethupathy, P. *et al.* Projections from neocortex mediate top-down control of memory retrieval. *Nature* **526**, 653–659 (2015).
6. Chettih, S.N. and Harvey, C.D. Single-neuron perturbations reveal feature-specific competition in V1. Nature 567, 334-340 (2019).
7. DeNardo LA, Liu CD, Allen WE, et al. Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nat Neurosci.* 22(3):460-469 (2019).
8. Ye L, Allen WE, Thompson KR, et al. Wiring and Molecular Features of Prefrontal Ensembles Representing Distinct Experiences. *Cell.* 165(7):1776-1788 (2016).
9. Wiltgen BJ, Zhou M, Cai Y, Balaji J, Karlsson MG, Parivash SN, Li W, Silva AJ. The hippocampus plays a selective role in the retrieval of detailed contextual memories. Curr Biol. 20(15):1336-44 (2010).
10. Sahay A, Scobie KN, Hill AS, et al. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature.* 472(7344):466-470 (2011).
11. O'Connor DH, Hires SA, Guo ZV, et al. Neural coding during active somatosensation revealed using illusory touch. *Nat Neurosci.* 16(7):958-965 (2013).
12. Cossell, L. et al. Functional organization of excitatory synaptic strength in primary visual cortex. Nature 518, 399–403 (2015).
13. Ko, H. et al. Functional specificity of local synaptic connections in neocortical networks. Nature 473, 87–91 (2011).

nature portfolio

Reviewer Reports on the First Revision:

Referees' comments:

Referee #2 (Remarks to the Author):

The exciting and compelling studies presented in the manuscript by Yadav et al address an old and key problem in memory: how context representations are formed in the brain. The authors use a number of state-of-the-art approaches to show how the encoding of complex, conjunctive contextual memories are driven by representations of their constituent features. For example, they show that the anterior cingulate cortex (ACC), but not the entorhinal cortex, drives contextual ensembles in the CA1 to modulate memory retrieval. Importantly, the authors determine that while the hippocampus codes for a complex, conjunctive representation of context, its component features are encoded by the AC, thus identifying a parallel but qualitatively distinct memory code from that in the CA1. The manuscript is logically structured, well written, and in the revised version the authors addressed most of the concerns expressed by the reviewers with new experiments and analyses. This is an impressive paper that will have an impact in a very active area of memory research.

Referee #3 (Remarks to the Author):

This is a very impressive manuscript that uses a wide range of methods to demonstrate the importance of ACC in contextual feature representation and retrieval. It also involves a technical tour-de-force and uses a number of innovative approaches to demonstrate the phenomenon. A particularly laudatory aspect is the careful attention to behavior; this is in addition to the use of state-of-the-art technical approaches. The analysis is generally done quite well. Overall, this will make a very nice contribution to the literature.

I don’t believe that additional experiments are required given all that the authors have done; clarifications and potentially additional analysis will be helpful. In general, given how much is done and presented, it can appear that each experiment is not as fully presented. The comments below will hopefully help future readers.

1. It is stated that about 10% of neurons were sensitive to context. The listed method is prone to false positive detection; the authors appear to account for this but the approach is opaque as currently described. A 1 std threshold for the mean ITI seems quite low and prone to false detection. The 40% threshold use is unclear even after reading Ext Fig 2. Please provide a more detailed explanation for this figure/method. Also, there should be a correction for multiple comparisons (i.e., a $p<0.05$ can lead to false positives when multiple tests are conducted).
2. Figure 1h show PC projections. Are these trial averaged or single trials? Was only a condition used to make the PC or all conditions were in the average? What is the VAF for the PCA for the trial averaged approach. I don’t feel strongly about this, but authors can consider use of demixed PCA as it can account for condition invariant and condition dependent variance.
3. “In fact, almost all neurons…significant and prominent responses …” What test was used?

“However, state-space analysis … no significant separation” What test was used?
There are multiple missing statements of test used, DOF, etc.
4. It is not clear what we are supposed to glean from Fig 2e, f? The traces are compressed and I can only vaguely detect a response or change. Consider presenting in an alternate manner.
5. Figure 2g,h is very interesting. However, it is not appropriate to lump sessions and animals. This requires a linear mixed effects model to account for sessions per animal.
6. Results of Fig 4c are important. The cumulative plot is compelling for the single animal. It is unclear what ”quantification of difference of mean onset time of AC and CA1 during training and retrieval, as determined by fitting onset curve to an exponential function.” What does ‘mean onset’ represent for an exponential curve?
I also believe three animals were used here? Does 10f show the data for the other two mice?

7. For the selective optogenetic inhibition experiments, authors performed comparisons between TRAP-Feature vs. TRAP-Habituation, and Light-on vs. Light-off in TRAP-stGtACR2 group. If will be more persuasive if the author can add one more comparison experiment between TRAP-stGtACR2 group and TRAP-mCherry/tdTomato group.

8. For the selective optogenetic inhibition experiments, according to the FosTRAP2 related reference, the TRAP system starts to occur within a 6-hour window centered around the 4-OHT injection. Given that, is there any specific reasons for the authors to perform inhibition 5 days after Day1 of retrieval? Do the authors also have the behavior data with/without inhibition during Day2 to Day5 of retrieval?
9. For the cFos+ staining and overlap counting in ExFig7, more details are required on how the cells were quantified. How many sections were used per animals? Were sections consecutive or spaced, and if spaced, how much? For each data point in ExFig7e, was it counted across the whole session or averaged across multiple fields of views?

Minor
1. The two alternatives in paragraph 2 could be more clear. It took a couple of readings to perhaps distinguish the difference.
2. “Of note, since some mice did not reliable classify neutral context “ - should be ‘reliably’
3. For the retrograde tracing in Fig 2a (and ExFig 5a), the authors suspected extra-hippocampal brain circuits provide feature inputs to CA1 and tried to catalogue key brain regions that send direct projections to CA1. However, the authors targeted the whole dorsal hippocampus and injected AAVrg-CAG-tdT to both CA1 and CA3, which might cause “virus leaking” and expressing across the whole area and might lead to over-retrograde labeling. An injection in CA1 only might give a more persuasive retrograde tracing result. In any case, this should be discussed.
4. For the inhibition-imaging experiments, Fig 3b, ExFig 5b and Method-Viral injections session, authors claimed to use AAV1-CaMKII-stGtACR2 (Addgene, Cat No. 105669-AAV1) for inhibition-imaging experiments. However, in Method- Paired Optogenetic Inhibition & Two-photon Imaging Session, author claimed to use AAV9-CaMKII-stGtACR2. Is there a typo in AAV serotype mentioned above?
5. For the optogenetic inhibition experiments, in main manuscript and Method, authors claimed to use CaMKII-eYFP (Addgene Cat No. 105622-AAV1) as control group. However, in caption of Fig 2h-i,

authors claimed ‘’opsin (GtACR) vs. control (mCherry) cohorts’’. What is the actual fluorophore for the control group?
6. In addition, possible typo in Methods-Anatomical Tracings session: AAVrg-CAG-tdT rather than AAVrd-CAG-tdT
7. For the inhibition-imaging experiments, authors injected soma-localized AAV1-CaMKII-stGtACR2 to LEC and ACC and performed circuit-specific inhibition locally. Although both ACC and LEC do send strong projections to ipsilateral CA1, they also have projections to some other brain regions. Soma manipulation at LEC or ACC may also inhibit their inputs to other downstream areas and cause deficits in other functions. This should be discussed.
8. The Kitamura Science 2017 paper seems relevant and worth citing and reference.

Referee #4 (Remarks to the Author):

The manuscript by Yadav et al. addresses the question of what inputs to the dorsal hippocampus represent the individual features that together define contexts which are robustly represented in the hippocampus? The major findings are that:
1. in hippocampus most neurons encode conjunctive representations of context, i.e., encode all the features of one context similarly
2. inhibition of anterior cingulate cortex (AC) but not lateral entorhinal cortex (LEC) inputs disrupts performance in a recall task which uses features instead of the entire context
3. AC but not LEC inhibition also disrupts the activity of context-encoding hippocampal ensembles
4. the AC contains a higher proportion of cells which encode individual features but not the entire context, and a lower fraction of context encoding cells than hippocampus
5. inhibiting TRAP’ed neurons during feature-based discrimination in the AC is sufficient to disrupt subsequent feature-based discrimination
6. during training, context/feature specific neurons in hippocampus are recruited before those in AC whereas during retrieval this order is flipped
7. The activity of feature selective AC neurons is correlated with the corresponding context selective hippocampal neurons, and overall aversive features are overrepresented among neurons which are strongly correlated
8. Initially shock responsive neurons eventually acquire selectivity for the aversive context (hippocampus) and associated features (AC)

Overall this is an interesting and well done study. I think the general question of how different brain regions work together to bind feature encoding into representations of context is interesting and important. That being said, there are several places where what is being shown is not fully clear, and some additional detail / clarification is necessary. More importantly – what is exciting about this paper is showing distinct roles for the AC and hippocampus. While I think most aspects of this have been done well, there is still one major gap, which is showing that hippocampus -> AC communication is important during training but not during retrieval (see comment #6 below). I realize this paper has already gone through one round of review. That being said some experiments inhibiting the hippocampus while imaging from the PFC would solidify the distinct but complementary roles for these two structures that the current paper implies, but does not directly show. In particular, I think the big conceptual insight here is showing how these two structures work

together, so showing that the hippocampus is not required for feature representations in the AC (whereas the AC is required for context representations in the hippocampus) would, in my opinion, be important to fully establish the key point of this manuscript. Again, because I am in the unusual position of entering the review process partway through, I will defer to the editors about this.

Specific Comments:
1. In Figure 4c, it's not totally clear how the population of neurons are being defined. Are these feature selective neurons in AC and context selective neurons in hippocampus? If so, this should be stated more clearly, and similar curves should be shown for context selective neurons in both structures and feature selective neurons in both structures. I.e., is this change in temporal relationships a general property of AC vs. hippocampal activity or a general property of feature vs. context representations, or is it specific for AC feature – hippocampal context representations?
2. In Figure 4c, the K-S test comparing distributions across neurons should not be the basis for establishing significance – this is simply too sensitive of a comparison and it's hard to believe that this would not show significant differences for almost any populations of neurons. There should be some comparison at the level of mice, e.g., a paired comparison of the mean time of recruitment for AC feature vs. hippocampal context neurons for each mouse.
3. In Extended Data 9f, there should be some quantification / summary to indicate the exact numbers / fractions of neurons that are labeled using TRAP during habituation vs. feature discrimination.
4. It's not clear that the TRAP'ed neurons are really feature-selective neurons as is implicitly assumed – unless the authors perform imaging of the TRAP'ed neurons to confirm that they are encoding features, they should be more circumspect about the language related to this.
5. The statement "Thus multiplexed single neuron activity robust population level representations of features" doesn't totally make sense to me. How exactly was this assessed / what specific comparisons support this?
6. The final statement "Such salient representations are further supported by bi-directional hippocampal-neocortical interactions, where hippocampus instructs salient feature representations in cortex during training, and cortex performs targeted recruitment of hippocampal representations during retrieval" – the latter part of this is supported by this study but the first part is mostly speculation. There is no direct evidence that the hippocampus instructs feature representations in AC during training. Showing this, by inhibiting the hippocampus while recording from the AC would be very powerful. Similarly, there is an implicit assumption that the direction of communication is purely AC->hippocampus during retrieval, but this has not been shown directly – inhibiting the hippocampus and showing that this does not affect AC feature representations during retrieval would accomplish this.