Peer Review File

Manuscript Title: Fast odour dynamics are encoded in the olfactory system and guide behaviour

Editorial Notes:

Please note that a third referee (Referee #1) was initially asked to evaluate this paper but was unable to return their report.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #2 (Remarks to the Author):

The manuscript by Ackels and co-workers presents a set of experiments showing that mice can discriminate correlated from uncorrelated changes in odorant concentrations fluctuating at frequencies higher than sniff frequency and provide data showing that neural activity of mitral/tufted (M/T) cells in the olfactory bulb can be used to decode whether the two odorants presented are correlated in time. This is an important finding that validates for the first time a speculation by Hopfield who suggested that the turbulent nature of odor transport allows fluctuations in odor concentration to convey information on odor source. I find the studies well-designed and innovative. My enthusiasm is diminished by problems with the presentation of the data and the lack of statistical analysis supporting a subset of the claims. I do not think these are unsurmountable problems. Below I raise several points to improve presentation of the data.

Major Comments

1. The fact that the mice are able to detect high frequency fluctuations at frequencies above the sniff rate is counterintuitive because it is generally assumed that diffusion of odorants in the mucus and the response of olfactory sensory neurons (OSNs) to odorants act as a low pass filter that would abolish information in high frequency fluctuations. Indeed, there is almost no literature on high frequency responses of OSNs to odorants. For example Ghatpande and Reisert (J Physiol 589.9 (2011) pp 2261–2273) mention that they attempted 10 Hz stimulation, but do not show the data because they did not find a relationship of action potential firing to stimulation (likely because of problems with stimulus control). I think that the approach that Ackels et al took by providing a proof of concept model and recording calcium changes in the glomerular OSN input is reasonable, although it is limited by the slow responsiveness of GCaMP and the bulk measurement of glomerular activity. A limitation was that there was no information on correlated vs uncorrelated odorant in the glomerular recordings (Extended Data Fig. 8). For the response of glomerular OSN input to correlated vs. anti-correlated stimuli did the authors attempt to decode the stimulus using deep network algorithms (https://github.com/KordingLab/Neural_Decoding)? For the OSN input imaging one would expect a larger difference in responses for glomeruli that respond differentially to the two stimuli. Were the glomeruli tested for their differential response to the two stimuli? How many glomeruli were in the field of view? Please show an image of the field of view used for glomerular imaging. When the classifier was trained with small numbers of glomeruli was there improved accuracy for a subset of the tests? Was there no information on stimulus correlation even when the frequency was dropped to 2 Hz? Was a smaller window size tested for decoding (say 20-50 msec). It is disappointing that the results are not shown for a large number of glomeruli.

2. The fact that evidence on transmission of information for high frequency odorant fluctuations from OSNs to the olfactory bulb is limited makes evidence on behavioral discrimination key for this manuscript. In my opinion the authors have made a thorough attempt at making sure that the responses of the mice in the AutonoMouse setup are valid as shown in Extended Data Fig. 5. However, the manuscript would benefit from a more thorough exclusion of the possibility that the two other senses that are exquisitely tuned to detecting high frequency oscillations (the auditory and somatosensory systems) are not mediating the behavioral responses. For the somatosensory system the authors provide evidence that a threshold analysis of the air flow does not yield accurate recognition of correlated vs. uncorrelated odorant. However, please provide a Fourier transform analysis of the air flow signals and decoding of the stimulus from air flow. Also, did the authors run a control where the whiskers of the mice were trimmed? For the auditory system did the authors place a microphone in the AutoMouse? Please record the sound for the different trials and provide Fourier transform/decoding analysis of the data. Also, if necessary was a noise maker used to mask the information provided by sound?

3. For the time-course for decoding analysis (e.g. Fig. 1j, Fig. 3g) it appears that decoding could differ from shuffled for early time points. This would be useful information. Please provide analysis of early time points with small time windows.

4. The evidence for differential responses of M/T cells to odorant fluctuations is key. Unfortunately, the presentation of the data is confusing and the data analysis and discussion of the relationship to published data could be improved. First, the presentation of the data are confusing. The line for OSN decoding cannot be discerned in Fig. 3f and inclusion of both OSN terminal and M/T cell data in Extended Fig. 8 is confusing. Extended Fig. 8 should be split into two figures (OSN input vs M/T responses) and the figure legend should be edited to provide a better explanation of the relevant information. Furthermore, the analysis of extracellular recording can be improved. The example unit in Fig. gii appears to clearly differentiate between correlated and uncorrelated. In fact, the response of that unit resembles responses of single M/T units found to respond to 10-200 msec duration of optogenetic activation of glomerular input by Li et. al. J. Neurosci., November 26, 2014 • 34(48):16058 –16064. Are these data concordant with the data of Li et al? Furthermore, the criterion for differential response of single units is not stated. Decoding analysis for low unit numbers is not provided. Also, the PCA analysis could be useful, but as presented contributes little to the validation of the M/T responses. Why is the accuracy calculated through PCA analysis starkly different to that of classifiers trained on summed spikes with small windows? Finally, the whole cell data in Extended Fig. 8 appear to provide evidence for differential responses to correlated vs. anticorrelated input. Could a subset of these M/T cells be classified as cells that respond differentially to the two stimuli?

5. The section on the method for generation of odor plumes from nearby and distant points is succinct and does not provide enough information for the reader to understand the measurement. Please provide diagrams of the indoors and outdoors setups. Given that there is a full speed fan drawing air into the PID will the indoors setup reproduce a situation that an animal is likely to encounter? Finally, please do not state that these conditions are "turbulent" unless this statement is backed by the physics underlying turbulent flow. "Complex" flow patterns may be more suitable. Finally, is the outdoors correlation significantly different from one for the one source case? Did this differ for different volatiles?

6. A substantial number of statements are not backed up by statistical analysis. For example the claim that mice perform above chance for probe trials in Figs. 4j,k is not backed up by a statistical test. For all figures where there is a comparison please show statistical differences with an asterisk and in the legend provide all the information for the statistical test used. For example, for Fig. 4di are some correlations statistically different? Which multivatiate test was used? What was the sample size? Are the data distributed normally? Are interquartiles shown (all that is stated in the legend is that what is shown is mean+SD)? Please revise the statistics through the document.

Minor comments

1. Line 44 change “prowess” to “powers”

2. Line 54, this reference should be to Fig. 1c

3. Line 133, please break into two paragraphs. Please add blank line between paragraphs.

4. The example of the ROI in Fig. 3c appears to be a dendrite. Yet, in the methods it is stated that all measurements were from somata. In fact, the information content may differ between somata and the dendrites within the glomerulus. Did the authors perform evaluation of dendritic responses for the M/T Cells?

5. Line 155: “Consistent with the Ca2+ imaging results, single-units did not follow high-frequency stimuli, but the average firing rate of 19% of units (18/97) again differed between correlated and anti-correlated stimuli (Fig. 3m,n and Extended Data Fig. 8g,h).” Please modify the sentence. First: Ca2+ imaging does not follow high-frequency stimuli because of technical issues. Second: What is the basis for classifying these as units that display a statistically significant difference in firing.

6. Line 393: “o, Average accuracies for sets of linear classifiers trained with different rolling window sizes (colour-coded), when the number of units available for the classifier was varied (n = 97, mean +/- SEM from 6 individual animals; see also Methods and Extended Data Fig. 8).” This figure shows the dependence of accuracy on window size as opposed to the number of units. Please show separate graphs for window size and number of units.

7. It is surprising that for the awake recordings in Extended Data Fig. 8 b the animal delays breathing for nearly one second. This reminds me of the response to irritant stimuli. Could it be that the concentration of the stimuli is so high that it is being detected as an irritant? Did the awake mice slow their breathing when presented with the odorants?

8. Please state how many animals were tested for each measurement.

9. Line 141. How was the significant difference determined for the integral responses? It seems integrating the response would loose temporal information. Were the responses subtracted before calculating the delta response integral?

10. Fig. 3n provides no information on differential responses. Please provide a raster of the delta PSTH.

11. The time scale bar is missing in Fig. 4g.

12. Line 198. On what basis is this statement made: “Notably, they reliably categorized correlated stimuli as “same source” and uncorrelated as “source separated””

13. Line 197. “that were derived from turbulence measurements”. Were these really measurements of turbulence?

Referee #3 (Remarks to the Author):

Schaefer and colleagues tackle an exciting, important and underexplored issue in sensory

neuroscience, specifically olfaction. It is often not appreciated that chemical stimuli in natural conditions arrive at sensors in fluctuating, intermittent manner due to the inevitable physics of the world. These fluctuations have a lot of high frequency components that can be dismissed as noise that needs to be averaged out. However, could animals make use of these fluctuations to infer important things about the olfactory world? In this paper, the authors offer evidence that mice can make use of high frequency olfactory information to make decisions. Related, the study also offers a demonstration of the idea that decorrelation across time can be exploited behaviorally (by mice). In addition, the authors also offer some support for neural coding of high frequency information in olfactory stimuli. Overall, this is an intriguing study that is both creative and novel. The behavioral work is rigorous and convincing, but I didn't find the neural recordings all that convincing (but it doesn't at all detract from the importance of the behavior – so not a problem!). I elaborate below with specific comments.

Abstract, line 6: It seems like "as little as" should be followed by a single number, rather than 10 or 25 ms.

Line 52/54. This refers to Fig 1b

Line 61-67 / Extended data Fig2: The authors show that a simple biophysical model is sufficient to have neurons exhibiting different responses to the two types of pulses despite the lack of faithful stimulus-following in time. For completeness and to be even more convincing they should add a simple decoding analysis similar to the one performed for OSN responses (fig 1i,j) showing decoding as a function of network size. I also wonder if adaptation could play some role, but perhaps worrying about all the details is not helpful!

Line 66: the logic here ignores concentration fluctuations - that is small differences of responses can also be interpreted as different concentrations. How will the brain know which is the cause of the different amplitude responses?

Line 77: I was unable to find in the methods whether classification is on single trials (and not trial-averaged). I may have overlooked this, but please clarify.

Line 81: As far as I can understand, timing and concentration are not really independently coded. If you were given two traces that are slightly different, you couldn't tell if the difference is due to timing or concentration of odor. The confusion matrix should be across modality of decoding.

Line 84 or so: The authors should state in the main text that the task is a go/no go design (line 84). They use as accuracy the mean of hits and correct rejects. Due to the asymmetry of the go/no-go paradigm, it would be best to show a version a plot of Fig 2k and Fig 4k separately for S+ and S- in the supplement. This would show that the increases in accuracy is properly distributed across the two options. (for example, mice could be mostly licking more or mostly licking less, with a very biased response hence the non-saturating accuracy)

Line 89, figure 2c: I really worry that this result (that mice simply cannot detect flickering stimuli very well) is telling us something. The authors argue that this is a poor test of temporal abilities and go on with their correlated-anticorrelated stimuli for the rest of the paper. I applaud the rigor and the various controls the authors have done – an excellent scientific process. But it is indeed puzzling why mice can't separate out, for example, 6 Hz vs 20Hz stimuli, very well. I wonder if the authors can discuss this further, along these lines (for example):

If you correlate two traces with the same phase but different frequencies (as in stimuli used for Fig 2c), you'll get an envelope rather than a zero-mean correlation – this could make it harder for mice to detect. Maybe the authors could show the temporal correlation for those stimuli? The stimuli the authors are using later (fig 2d, for example) for the correlated-anticorrelated expts are at one end of the spectrum (if you sum the stimuli, ignoring identity, the pulses go from 0 to 2 for

correlated and stable at 1 for anti-correlated – could mice be using some summed intensity metric?). If you do the same calculation for the 6hz-20hz, the envelope you get is somewhere in between – the authors could show those, and that might partly explain why mice are worse in Fig 2c?

Lines 145-147: It would be good to show that non-linear classifiers, or some metric of information, can pull out the differences from glomerular data. M/T cells are not pulling out information out of nothing!

Figure 3: All the physiology seems to be done in rhythmically breathing animals, where things are more predictable. As seen in sniffing traces during behavior, there is natural fluctuation in breathing that may make things harder to decipher.

Line 169: I'm not sure if one can actually ATTRIBUTE the location (that happens when the agent gets closer). It's more that one can say they are separate sources (or the same source, spatially).

Line 191: once again - the information is NOT about the location, it's more about single or multiple sources

Line 195: It looks like mice take a long time to learn... does this mean that this is actually hard to do, and not very natural for mice?

Author Rebuttals to Initial Comments (please note that the authors have quoted the reviewers in blue text and responded in black text):

Referee #2 (Remarks to the Author):

The manuscript by Ackels and co-workers presents a set of experiments showing that mice can discriminate correlated from uncorrelated changes in odorant concentrations fluctuating at frequencies higher than sniff frequency and provide data showing that neuralactivity of mitral/tufted (M/T) cells in the olfactory bulb can be used to decode whether thetwo odorants presented are correlated in time. This is an important finding that validates for the first time a speculation by Hopfield who suggested that the turbulent nature of odortransport allows fluctuations in odor concentration to convey information on odor source. Ifind the studies well-designed and innovative. My enthusiasm is diminished by problems with the presentation of the data and the lack of statistical analysis supporting a subset of the claims. I do not think these are unsurmountable problems. Below I raise several points to improve presentation of the data.

Thank you for the encouragement and the constructive comments! We have addressed these with new experiments, in particular additional imaging and electrophysiological recordings, improved the presentation of the data throughout and substantially augmented the statistical analysis as suggested.

Major Comments

1. The fact that the mice are able to detect high frequency fluctuations at frequencies above the sniff rate is counterintuitive because it is generally assumed that diffusion of odorantsin the

mucus and the response of olfactory sensory neurons (OSNs) to odorants act as a lowpass filter that would abolish information in high frequency fluctuations. Indeed, there isalmost no literature on high frequency responses of OSNs to odorants. For exampleGhatpande and Reisert (J Physiol 589.9 (2011) pp 2261–2273) mention that they attempted10 Hz stimulation, but do not show the data because they did not find a relationship of action potential firing to stimulation (likely because of problems with stimulus control). I thinkthat the approach that Ackels et al took by providing a proof of concept model and recordingcalcium changes in the glomerular OSN input is reasonable, although it is limited by theslow responsiveness of GCaMP and the bulk measurement of glomerular activity. A limitation was that there was no information on correlated vs uncorrelated odorant in theglomerular recordings (Extended Data Fig. 8). For the response of glomerular OSN input tocorrelated vs. anti-correlated stimuli did the authors attempt to decode the stimulus usingdeep network algorithms (https://github.com/KordingLab/Neural_Decoding)?

Thank you for this suggestion (that was similarly made by reviewer 3).

We have now performed more glomerular imaging experiments to get a richer dataset toanalyse and assess what information is available in the OSN Ca^{2+} signal. We recorded responses to 3 different odour pairs at frequencies (as suggested below) of 2 Hz as well as 20 Hz. With now a total of 435 glomeruli-odour pairs (3 odour pairs, 145 glomeruli recorded across 5 animals) we find – using the same classifier analysis as before – that2 Hz correlated vs. anti-correlated stimuli can indeed be classified with up to 70% accuracy **(Figure R1a)**. For 20 Hz stimuli classifier accuracy of ≈57% can be reached with 100+ glomeruli-odour pairs **(Figure R1c)**. Following the reviewer's advice, we thenwent on to expand on the classifiers we had used on the new data (with 3 odour pairs andmore repeats). To investigate the responses in more detail we trained directly on glomeruli and separately for each odour pair, instead of using glomeruli-odour pairs as in the previous analysis. In order to obtain interpretable results, we imposed a sparsenessconstraint. Indeed, this approach readily allowed classification for the 2 Hz case (shown for one example odour pair in **Figure R1bi** and for all 3 odour pairs in direct comparison to the linear classifier without sparsity constraint in an appendix **Figure R21a,b**). Note that indeed – as expected if overfitting is not an issue – the shuffle control classification was robustly at chance **(Figure R1bii)**. Testing the same sparse linear classifier at 20 Hzrevealed accuracy that substantially exceeded chance for individual time bins (**Figure R1di** whilst shuffle controls remained at chance **Figure R1dii**).

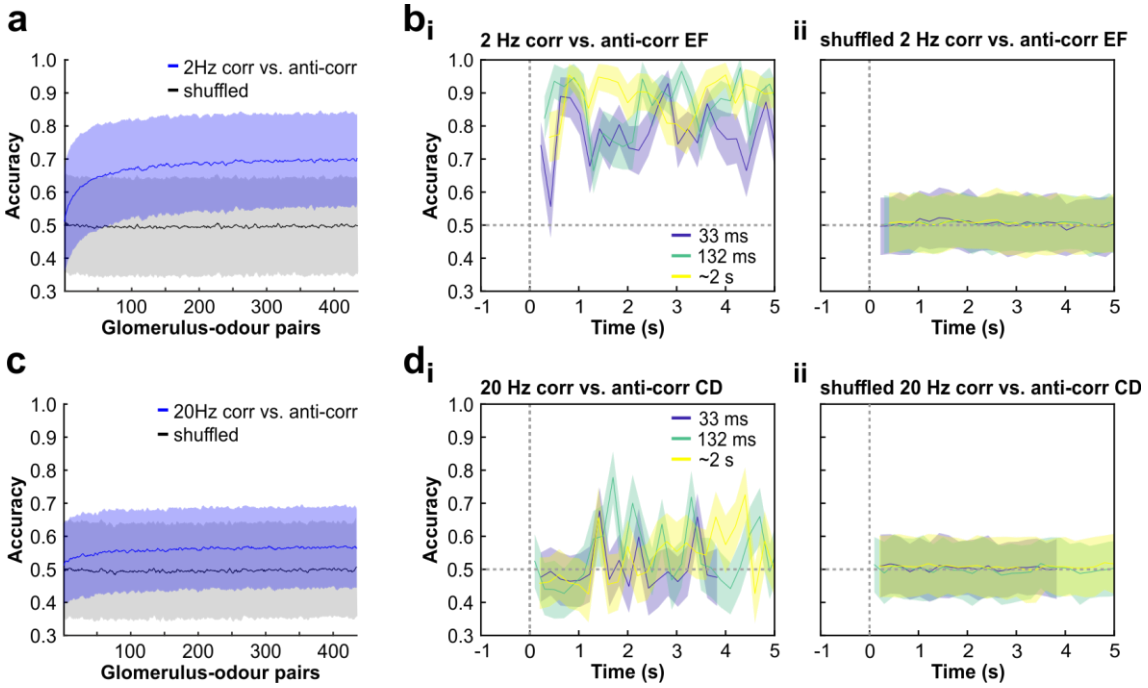

Figure R1: Classification analysis of glomerular imaging data. a, Classifier accuracy when trained over an increasing number of glomerulus-odour pairs (n = 435, 145 glomeruli from 5 individual awake animals and 3 odour pairs) when a linear classifier was trained on the summed 2 s response to correlated vs. anti-correlated stimuli at 2 Hz (black: shuffle control; mean ± SD of 500 repetitions). **b_i,** Accuracy of a linear classifier (with sparseness constraint) when trained on all responsive glomeruli (see Methods and Figure R21) in response to odour pair EF for 2 Hz correlated vs. anti-correlated (n = 145, from 5 individual animals) with a sliding window of different durations (colour-coded; mean ± SD for of repetitions) starting at 1 s before odour onset. Time points indicate latest extent of each window. **b_{ii},** Same as **b_i** with labels shuffled. **c,d**, same as **a,b** for 20 Hz correlated vs.anti-correlated odours. The patchiness observed late in the response in **d_{ii}** is because not all time points had responsive ROIs for all time windows (see also detailed Methods in the revised manuscript and Figure R21).

Classification, however, wasn't robust across time windows or time points or odours (seealso **Figure R21a** for all odour pairs). Thus, linear classifiers can indeed robustly reveal information about correlation structure at low frequencies and some information about correlation structure even at 20 Hz. This, however, is substantially less pronounced and robust than in the M/TC responses (**Figure 3** and **Extended Data Figure 9,10,11)**

Following the reviewer's suggestion, we then explored non-linear classifiers (summarized in the appendix **Figure R21c,d**). While state-of-the art deep learning networks might find structure that remains hidden otherwise, they derive complex, non-linear decision boundaries that are characterized by a large number of parameters. Therefore, they typically rely on large amounts of training data, prohibitive for the kind of experiments presented here. The small number of trials we have available relative to the number of glomeruli also means that the data are linearly separable, suggesting limited additional benefit from more sophisticated approaches. We thus decided to initially explore support vector machines with non-linear kernels because they are non- linear classifiers that can be parameterized with smaller numbers of parameters than deep learning approaches. Unfortunately, neither 3rd order polynomial nor radial basis function kernels substantially improved classification accuracy. In fact, both approachesconsistently resulted in below-chance classification accuracy on withheld training data for the 20 Hz case as well as for all shuffle controls **(Figure R21c,d)**, indicating potential overfitting. So, in short, while non-linear classifiers might be able to pull out information about stimulus correlations more effectively, the large number of parameters relative tothe number of trials available make training such classifiers very difficult. However, already with simply more data and more in-depth linear analyses we could indeed pull out information about correlation structure from glomerular imaging. We have now included this new data and new analysis in the new **Extended Data Figure 8** in the revised manuscript.
We should emphasize, however, that all this analysis is on slow Ca^{2+} imaging data, averaged across all OSN axons in a given glomerulus. The neural circuitry of the olfactorybulb of course has access to much more fine-grained information from OSN activity – action potentials with millisecond precision and thousands of individual fibres.

Finally, we want to highlight that – while the OSN imaging, possibly due to the low spatiotemporal resolution of the measurement technique as discussed above, allows us to extract only comparatively little information about whether stimuli are correlated or anti-correlated – the imaging and electrophysiology we show for M/TCs (**Figure 3**, new **Extended Data Figures 9-11**) does demonstrate that indeed high-frequency odourinformation readily reaches the OB. In fact, the simplistic pulse experiment with Ca^{2+} imaging recordings from OSNs in **Figure 1** supports this general notion as well. To furthercorroborate the point that neural activity in the OB is sensitive to high-frequency contentof odour stimuli, we have performed new experiments using similar high-frequency "blip" stimuli and measure M/TC responses with unit recordings **(Figure R2)**. Indeed, recordings from M/TCs readily allowed us to discern what kind of short odour pulse combination had been presented even when odour intervals were as little as 20 ms. We now emphasize this in the revised manuscript in the new **Extended Data Figure 10**.

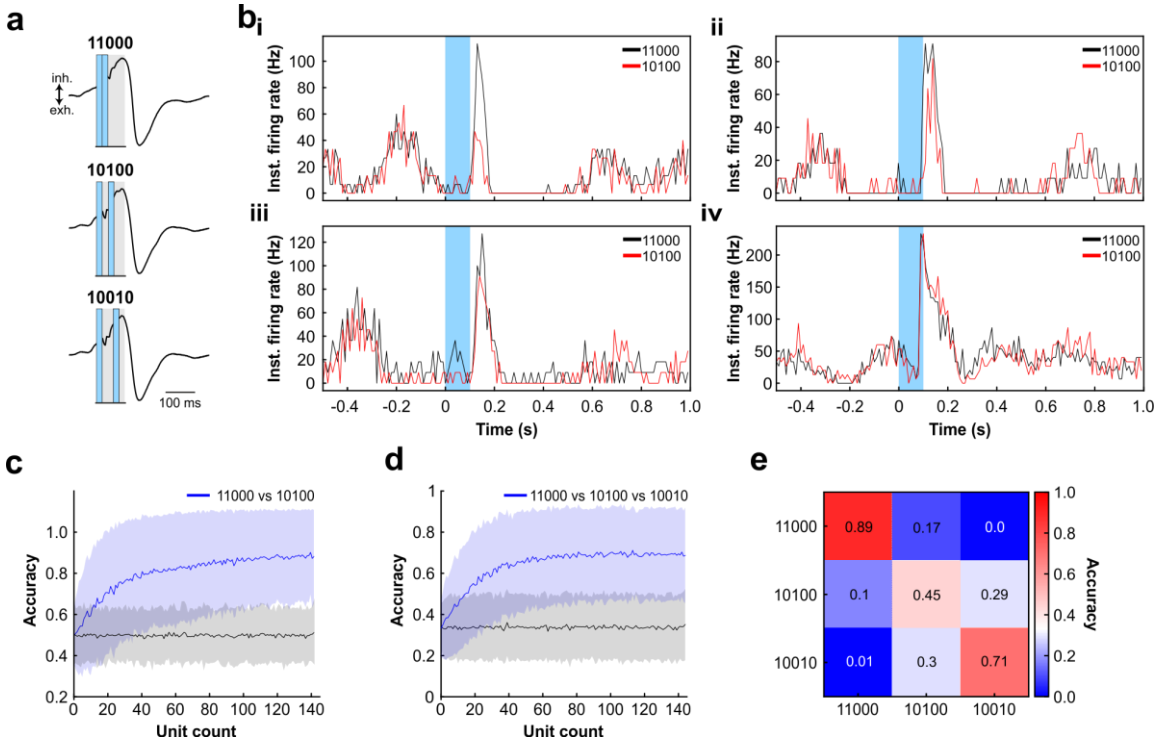

Figure R2: Unit recordings in response to short odour pulse combinations. a, Schematic of odour pulse stimuli timings in relation to the respiration cycle. Three combinations were presented, each trial 120 ms in length. For example, 11000 (top) consisted of a 40 ms odour pulse (light blue) followed by 80 ms of blank odourless air (grey); All trials were triggered at the onset of inhalation. **b, (i-iv)** PSTH (peri-stimulus time histograms) from four example units showing their average firing rate prior, during, and post odour stimuli (light blue vertical bar). Responses are either to 11000 (black) or 10100 odour presentation (red). The instantaneous firing rate was calculated for 10 ms windows. **c,** Accuracy of linear classifiers as a function of the number of units available for training/testing (mean ± SD; n = up to 145 units from 8 individual anaesthetised animals). Each classifier is trained on the summed spike count of the available units in a window of 500 ms starting at odour onset. The classifiers were trained on all but two trials, one 11000 and one 10100 trial, and the number of repeats between animals varied between 11 and 30. To account for this and to minimise the variability of the training set, trial number was bootstrapped to 1000 repeats. This was achieved by randomly selecting a repetition for each unit independently.The test set was isolated from the responses prior to bootstrapping and thus was not seen by the classifier until it was tested on it. Each classification was repeated 500 times with a different selection of units, and a different test set. The shuffled control (black) was accomplished by shuffling the training labels during each iteration of the classifier without shuffling test labels. **d,** Same as in **c** but classifying all three odour pulse combinations. **e,** Confusion matrix outlying the fractions that each trial type was classified as (n = 145 units from 8 individual animals). True labels are shown on the x axis and labels predicted by the classifier on the y axis. Accuracies correspond to maximum unit count shown in **c** and **d**. The classifiers can readily separate between trials containing a single 40 ms odour pulse. Accuracy is lower when distinguishing between an intermission of 20 or 40 ms but remains substantially above chance (chance = 0.33).

For the OSN input imaging one would expect a larger difference in responses for glomerulithat respond differentially to the two stimuli. Were the glomeruli tested for their differential response to the two stimuli?

This is an excellent point that will potentially help to mechanistically unravel how odour correlation structure is represented and might be processed. We have, therefore, now performed additional new experiments with a set of different odours, both individually as suggested by the reviewer and in correlated / anti-correlated combinations. **Figure R3**shows how the glomerular response to individual odours relates to their response to correlated / anti-correlated odours. Glomeruli that respond strongly to one but not the other odour have a selectivity of ~1 in **Figure R3h,i**; those glomeruli responding approximately equal to both have small selectivity values <0.5 in **Figure R3h,i**.

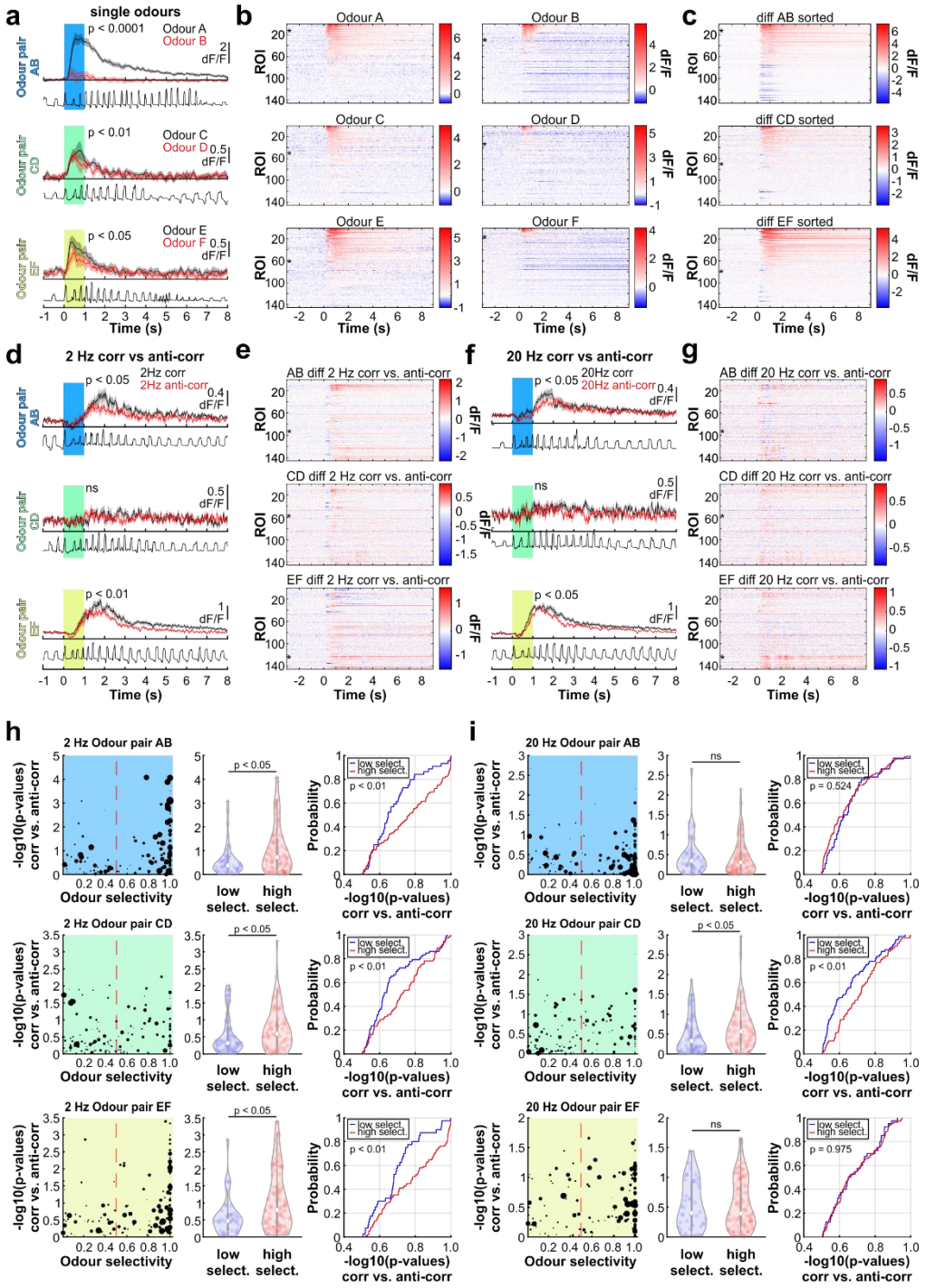

Figure R3: Glomerular imaging in response to single odours and odour correlation structure. a, Example glomerulus responses to the presentation of individual odours plotted pairwise (AB, CD, EF; mean of 6 trials ± SEM). Stimulation period (1 s) is indicated by vertical bar (blue, green and yellow). P-values are derived from unpaired t-tests comparing responses integrated over 2 s. **b,** Calcium transients plotted as colour maps sorted byresponse magnitude for individual odours. **c,** Difference between glomerulus responses to individual odours plotted pairwise as colour maps. Glomeruli are sorted by the average magnitude of response difference. **d,** Example glomerulus responses to presentation of correlated vs. anti-correlated odour pairs fluctuating at 2 Hz, mean of 12 trials ± SEM. P-values are derived from unpaired t-tests comparing responses to correlated vs. anti- correlated stimulation integrated over 2 s. **e,** Difference between glomerulus responses to 2 Hz correlated and anti-correlated odour stimulation plotted as colour maps, sorting as in **c**. **f-g,** Same as in **d-e** but for 20 Hz correlated vs. anti-correlated. **h,** Left: P-values derived from comparing trials of the summed 2 s response to correlated vs. anti-correlated odour stimulation at 2 Hz (unpaired t-tests) for three odour pairs (colour-coded) as a function of glomerulus selectivity to individual odours. Selectivity is calculated as the difference between the response to individual odours scaled by the summed response. A threshold is set at 0.5 defining glomeruli as lowor high selective. Dot size represents magnitude of the summed response. Middle: Comparison of p-values between low and high selective glomeruli (n =

145, $p < 0.05$, unpaired t-test). Right: Cumulative distribution function of p-values for low and high selective glomeruli ($n = 145$, $p < 0.01$, two-sample Kolmogorov-Smirnov test). **i,** Same as **h** for 20 Hz. Example glomerulus from **a,d,f** indicated with an asterisk in colour maps in **b,c,e,g**.

In both these groups we find individual glomeruli that show differential response to correlated and anti-correlated inputs (low p-values, i.e. high -log p in the figure). There was a tendency for selective glomeruli to be more likely to respond differentially. However, overall, we find no clear-cut relationship between odour selectivity and differential response to odour correlation structure. This indicates that factors other than odour selectivity, e.g. odour molecule interactions at olfactory receptors or feedback fromthe OB circuitry might play a role in shaping glomerular Ca^{2+} signals. We now summarizethese findings as part of the new **Extended Data figure 8** that describes OSN responses in more detail.

How many glomeruli were in the field of view? Please show an image of the field of view usedfor glomerular imaging.

We now provide detailed information on the number of glomeruli in a typical experiment(line 1475 of the revised manuscript) as well as example images for the experiments where the response to correlated / anti-correlated stimuli was measured in the OSNs in the new **Extended Data Figure 8** (replicated here as **Figure R4**) We also show a field ofview for the OSN double-pulse experiments in **Figure 1e** as well as in **Extended Data Figure 3a**.

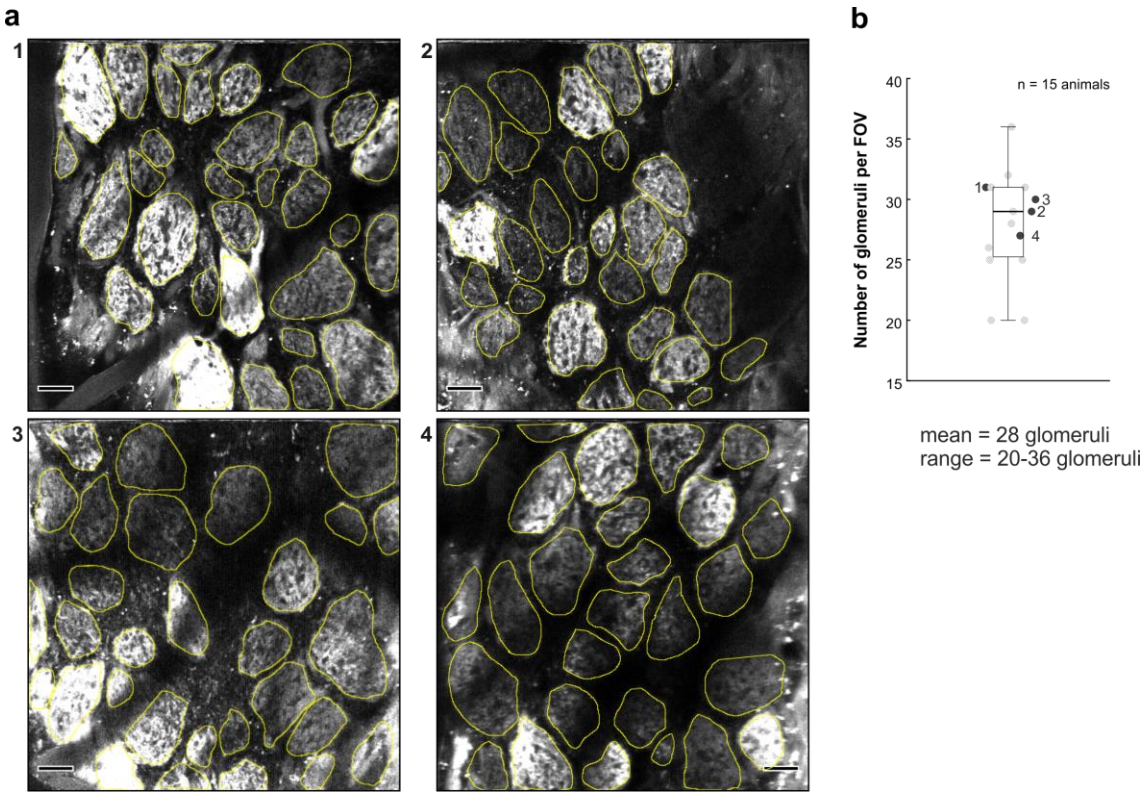

Figure R4: Example fields of view for glomerular imaging. a, Fields of view (FOV; 1-4) recorded from the dorsal bulb of individual animals. **b,** Number of individual glomeruli per FOV in all experimental mice (n = 15). The number of individually delineated glomeruli ranges from 20-36 with an average of 28 glomeruli per FOV. Labelled data points (1-4) correspond to FOVs shown in **a**. Scale bars: 50 µm. Edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers as described in the methods section.

When the classifier was trained with small numbers of glomeruli was there improved accuracy for a subset of the tests? Was there no information on stimulus correlation even when the frequency was dropped to 2 Hz?

We have now performed more experiments as described above as well as new experiments for 2 Hz stimulation **(Figure R1)**. Indeed, information about stimulus correlation at 2 Hz can be readily recovered even with simple linear classifiers from OSNimaging data.
We have now provided classifier accuracy as a function of the number of glomeruli in thenew **Extended Data Figure 8**. While for a smaller number of glomeruli stimulus classification in general works poorly for linear classifiers and 20 Hz stimuli, there are indeed individual glomeruli that show quite robust differences in response to correlated and anti-correlated stimuli. A classifier selectively trained on those glomeruli would indeed – despite the small number – perform well. We now show some of these examplesin the new **Extended Data Figure 8**, and replicated here **(Figure R5)**.

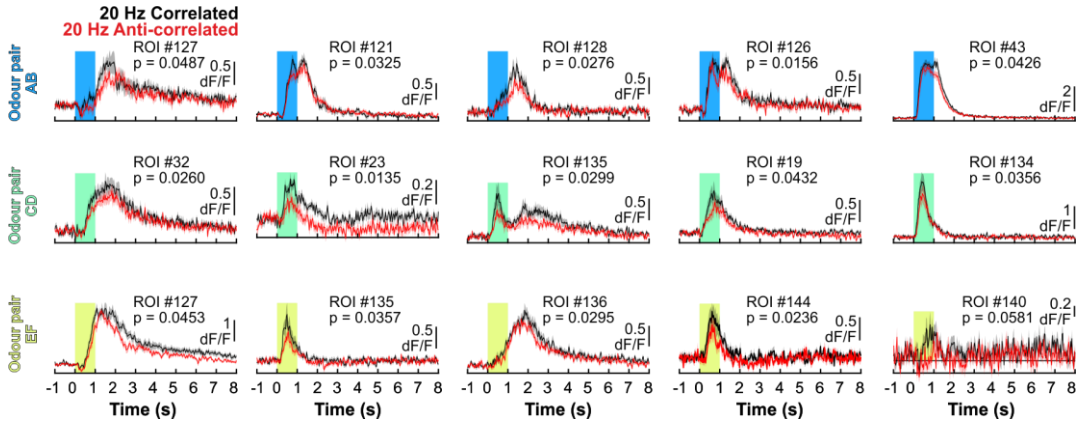

Figure R5: Glomerular responses to 20 Hz correlated and anti-correlated odour stimulation. Odourpairs presented either correlated (black) or anti-correlated (red) at 20 Hz are colour coded (AB, blue; CD, green; EF, yellow). The stimulation phase (1 second) is indicated by vertical bars. Responses of example glomeruli depicted here show significantly different activation based on the integrated response calculated over a 2 second window from odour onset (unpaired t-tests; AB: 6/145 (4.14%), CD: 14/145 (9.65%), EF:5/145 (3.45%)). Mean of 12 trials ± SEM, recorded from 5 individual awake animals.

Was a smaller window size tested for decoding (say 20-50 msec).

Thank you for pointing this out. We have now tested different decoding windows for theOSN imaging **(Figure R6)**. However, as Ca^{2+} imaging is relatively slow it is maybe unsurprising that shorter time windows do not provide substantially better classificationaccuracy. We comment on this in the revised **Figure 1j** and **Extended Data Figure 3**.

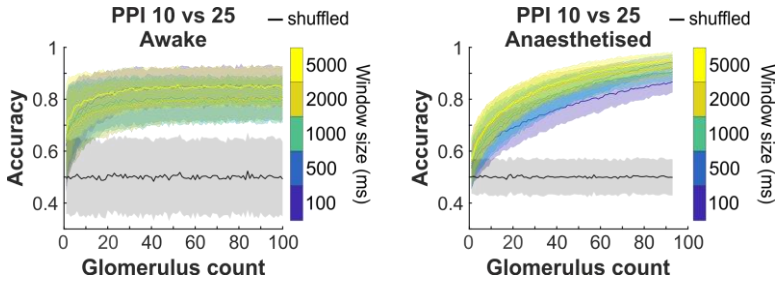

Figure R6: Linear classifiers trained with different response window sizes over increasing number of glomeruli. Classifier accuracy (mean ± SD) over an increasing number of glomeruli when a linear classifier was trained on the summed response to PPI 10 vs. 25 ms stimuli over different time windows (between 100 and 5000 ms, colour-coded; black: shuffle control; 500 repetitions). Number of glomeruli was n = 100 from 5 awake and n

= 93 from 4 anaesthetised animals.

It is disappointing that the results are not shown for a large number of glomeruli.

We have now performed additional experiments to increase the total number of glomeruli imaged to 145. We have also used additional different odour sets so that the total number of "glomeruli-odour pairs" was increased to 435. We have described the results of this in detail in the previous paragraphs and have incorporated this new data in the new **Extended Data Figure 8**.

2. The fact that evidence on transmission of information for high frequency odorant fluctuations from OSNs to the olfactory bulb is limited makes evidence on behavioraldiscrimination key for this manuscript. In my opinion the authors have made a thorough attempt at making sure that the responses of the mice in the AutonoMouse setup are valid as shown in Extended Data Fig. 5. However, the manuscript would benefit from a more thorough exclusion of the possibility that the two other senses that are exquisitely tuned todetecting high frequency oscillations (the auditory and somatosensory systems) are not mediating the behavioral responses. For the somatosensory system the authors provide evidence that a threshold analysis of the air flow does not yield accurate recognition of correlated vs. uncorrelated odorant. However, please provide a Fourier transform analysis of the air flow signals and decoding of the stimulus from air flow. Also, did the authors run a control where the whiskers of the mice were trimmed? For the auditory system did the authors place a microphone in the AutoMouse? Please record the sound for the different trials and provide Fourier transform/decoding analysis of the data. Also, if necessary was anoise maker used to mask the information provided by sound?

Thank you very much for these suggestions! We have now measured air flow and soundsmade by the valves for correlated and anti-correlated odour stimuli as shown in **Figure R7.** We placed a flow sensor (Honeywell AWM5101VN) and a microphone (RØDE NTG1)in close proximity to the AutonoMouse odour port and measured 286 trials in total for 2Hz (n = 75 correlated, n = 70 anti-correlated) and 40 Hz (n = 69 correlated, n = 72 anti- correlated). Spectral analysis shows distinct peaks not only for the modulation frequencybut also at higher frequencies corresponding to the "shattering" used to regulate concentration **(Figure R7)**. While there are distinct patterns, these do not differ betweencorrelated and anti-correlated stimuli. Moreover, even with more sensitive classifier analysis no differences between the correlated and anti-correlated trials can be detected**(Figure R7b,c,d,eiv)**. We have incorporated this new data in **Extended Data Figure 5d-g** and added a description of the experiments to the Methods (line 1228-1236). In a subsetof trials during training to increase diversity in stimuli we used more complex valve arrangements (combining several valves carrying the same odour) that resulted in different sound/flow patterns. Behaviour, however, showed little difference between these "complex" and the predominantly used "simple" trials **(Figure R8)**.

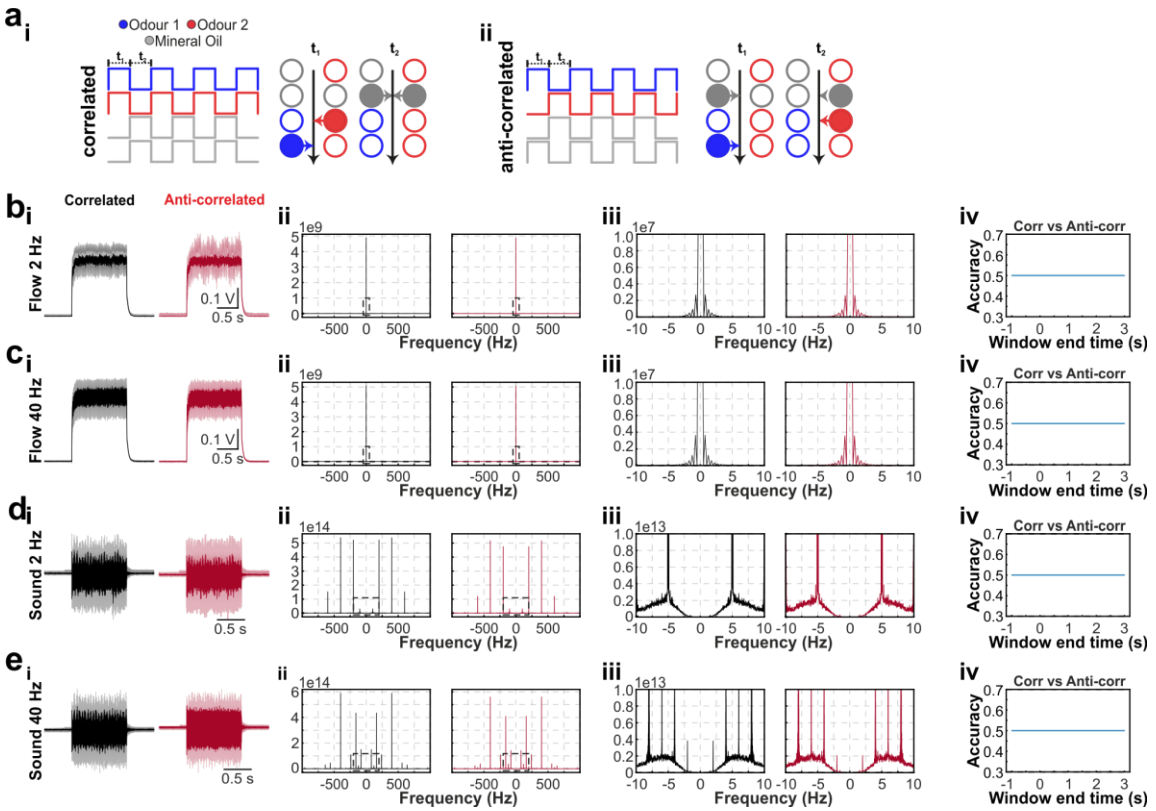

Figure R7: AutonoMouse flow and sound recordings. a, Detailed schematic of correlated (*left*) and anti- correlated (*right*) stimulus production; odour presentation (blue = odour 1, red = odour 2) is always offset by clean air (grey) valves at the same flow levels, to ensure that total flow during the stimulus is constant. t1 and t2 represent valve openings at the corresponding time points shown on the left. **a**$_{i}$, For a correlated stimulus, to produce the dual odour pulse at t1, odour 1 (blue) is delivered from one valve and odour 2 (red) from another valve. During t2 two valves contribute clean air. **a**$_{ii}$, For an anti-correlated stimulus, odour 1 and a clean air valve are open during t1 and odour 2 and a clean air valve during t2. **b**$_{i}$**,** Flow recordings of 2 Hz correlated (black, n = 75) and anti-correlated (red, n = 70) trials taken from the AutonoMouse odour port (mean ± SD). **b**$_{ii}$, Fourier transform of the flow signal from **b**$_{i}$, showing the power of the signal over a range of 1 kHz and **b**$_{iii}$, a zoom in over the range of 10 Hz indicated by the dotted box in **b**$_{ii}$. **b**$_{iv}$, Accuracy of a series of linear classifiers trained on an increasing window of the integrated signal starting from 1 s before trial onset shown in **b**$_{i}$. Classifiers were tested on two withheld trials, one correlated and one anti-correlated, and repeated 100 times. **c,** Same as **b** but for 40 Hz trials (n = 69 correlated and n = 72 anti-correlated). **d**$_{i}$**,** Audio recordings of 2 Hz stimuli using a microphone placed in close proximity to the AutonoMouse odour port. **d**$_{ii,iii}$, Fourier transforms of the audio signal from **d**$_{i}$. Note, whilst there are notable peaks at specific frequencies, these are present in both correlated and anti-correlated trials. **d**$_{iv}$, Accuracy of a series of linear classifiers as shown in **b** but using the modulus of the audio signal. **e,** Same as **d** but for 40 Hz. Note, whilst the sound profile and the Fourier transforms are different between 2 and 40 Hz, there is no difference detectable between correlated and anti-correlated trials.

Thus, air flow or valve noise do not predict whether a stimulus is correlated or anti- correlated. This is largely because (i) activity in odour-carrying valves is counter- balanced by non-odour-carrying valves so that in each condition the same overall noise pattern and flow emerges; and (2) we switch between different valve combinations during our entire training schedule. Thus, in order for noise to predict S+/S- trials, one would have to assume that specific combinations of valve noises would be learned by theanimal (i.e. the combination of valve1@pattern1 and valve2@pattern2 (together makingthe S+ stimulus) would sound different to valve1@pattern2 with valve2@pattern1 (S-) etc). In fact, the actual situation is even more challenging as a S+ stimulus is composed ofat least 4 valves (for the subset of “complex” stimuli 6 valves) simultaneously opening/closing (odour 1, odour 2 as well as two valves presenting the “odourless” mineral oil). Furthermore, for each trial a different combination of valves and/or

different random "shattering" ensures that no two trials carry the same sound pattern. Thus, an animal would have to learn and generalize across a large number of individual combinations of valve / flow patterns.
However, to test this very remote possibility, we introduced what we think are the most rigorous controls, "switch controls", where animals were presented with a new valve

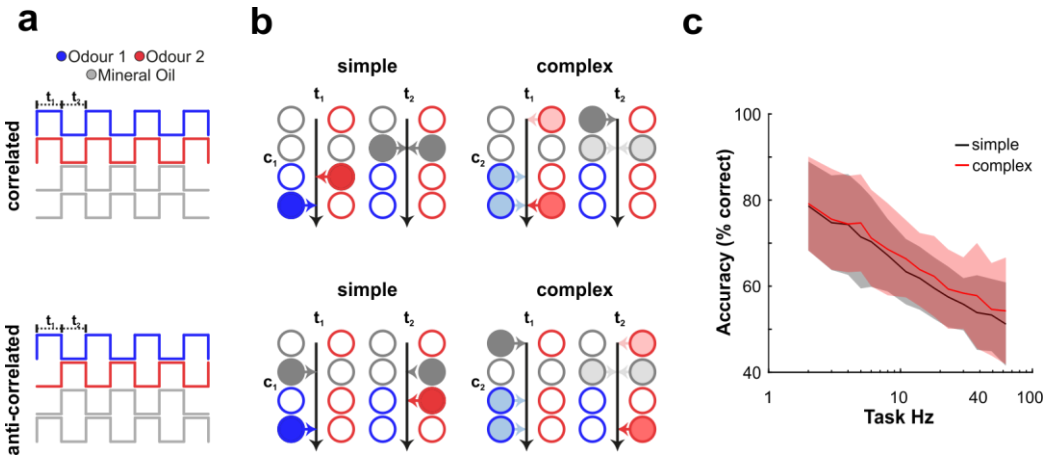

Figure R8: Correlation discrimination performance is unaffected by valve opening configuration. a, Detailed schematic of correlated (top) and anti-correlated (bottom) stimulus production; odour presentation (blue = odour 1, red = odour 2) is always offset by clean air (grey) valves at the same flow levels, to ensure that total flow during the stimulus is constant. Schematic of the use of valve subsets to produce the desired stimulus. t1 and t2 represent valve openings at the corresponding time points shown in a. **b,** c1 (**b**, left; simple) and c2 (**b**, right; complex) represent two possible configurations that could be used to produce the same resulting stimulus at the two time points. Opacity in the colours represents total concentration contribution to the resulting stimulus at the time point. For a correlated stimulus (top), to produce the dual odour pulse at t1, configuration c1 can be used where odour 1 (blue) is delivered from one valve and odour 2 (red) from another valve. During t2 two valves contribute clean air. Alternatively, configuration c2 can be used in which during t1 odour 1 (blue) is generated by 50% opening of two valves, with odour 2 (red) produced by 70% / 30% opening of two other valves respectively. **c,** Group accuracy performing correlation discrimination where stimulus pulse frequency is randomised from trial to trial, split into trials from simple (black) and complex (red) configurations of odour presentation (mean ± SD, n = 33 mice).

combination they hadn't encountered before (as valves and/or bottles had been moved to new positions). This new pattern – based on flow or sound – carries no information known to the animal about the nature (S+ or S-) of the stimulus as – depending on how bottles are filled with odours – the same valve pattern could result in correlated or anti- correlated odour 1/2 presentation **(Figure R9)**. Importantly, animals responded to trialsimmediately after switch exactly as they did before switch, indicating that any spurious non-olfactory correlation information that might have been there (although unlikely andwe could not measure any as discussed above) was not relevant for their decision making.

We realize that this description was very brief and difficult to understand in the original manuscript and – in addition to the new data from sound and flow measurements – we have now significantly expanded our description of the stimuli and the logic for our "switch" control in a revised **Extended Data Figure 5**.

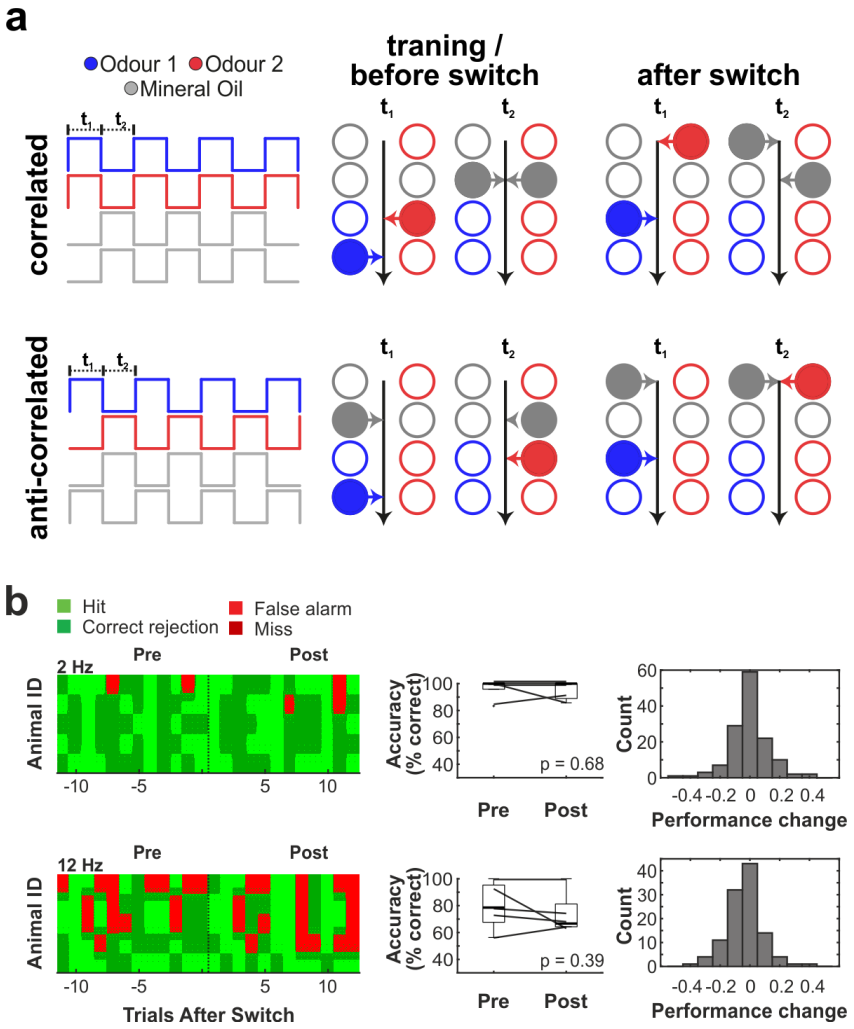

Figure R9: Valve switch control. a, Detailed schematic of correlated (top) and anti-correlated (bottom) stimulus production before (middle) and after (right) switching valves. For the switch control, a set of previously unused odour valves is introduced to rule out potential bias towards a specific valve combination when performing the odour correlation discrimination task. **b,** Left: Trial map of 5 representative animals during 2 Hz and 12 Hz correlation discrimination tasks before and after introduction of control valves (n = 12 trials pre-, n = 12 trials post- new valve introduction, new valve introduction indicated by black vertical dotted line. Each row corresponds to an animal, each column within the row represents a trial. Light green: hit, dark green: correct rejection, light red: false alarm, dark red: miss. Middle: Boxplots of mean accuracy for animals pre- and post-control (p = 0.68 for 2 Hz and p = 0.39 for 12 Hz, paired t-tests). Edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers. Right: Summary histograms of performance change for all animals during all "valve switch" control tests (see Methods) indicating that discrimination accuracy was based on intended olfactory cues. The five animals showing highest performance before the valve switch / bottle change (and thus the largest potential to drop in performance) were analysed.

Furthermore, we want to clarify an additional point: The OSN Ca^{2+} imaging experiments for correlated / anti-correlated stimuli indeed do not readily provide additional evidencefor access to high-frequency odour features (see, however, the discussion around **FigureR1**). However, as mentioned above, OSN imaging in response to odour "blips" (**Figure 1**)demonstrates access to temporal features on rapid time scales. Similarly, the imaging andelectrophysiology data from M/TCs in response to correlated / anti-correlated stimuli further demonstrate that the OB has access to such high-frequency components (**Figure 3**). The new experiments described in **(Figure R2),** where we measure the response of M/TC units in response to different combinations of 20 ms odour pulses, further corroborates this with yet another, new stimulus paradigm.

We now highlight these additional datasets in **Extended Data Figure 10** of the revised paper and have expanded our discussion about all the independent pieces of evidence weprovide for transmission of information about high-frequency odour fluctuations to the olfactory bulb (lines 218-222 of the revised manuscript).

3. For the time-course for decoding analysis (e.g. Fig. 1j, Fig. 3g) it appears that decoding could differ from shuffled for early time points. This would be useful information. Please provide analysis of early time points with small time windows.

We have now performed a new decoding analysis with more detail for earlier time points**(Figure R10)**. This is summarized and included in the new **Extended Data Figure 10** aswell as a new panel in **Extended Data Figure 3d** and main **Figure 1j and Figure 3g** of the revised manuscript.

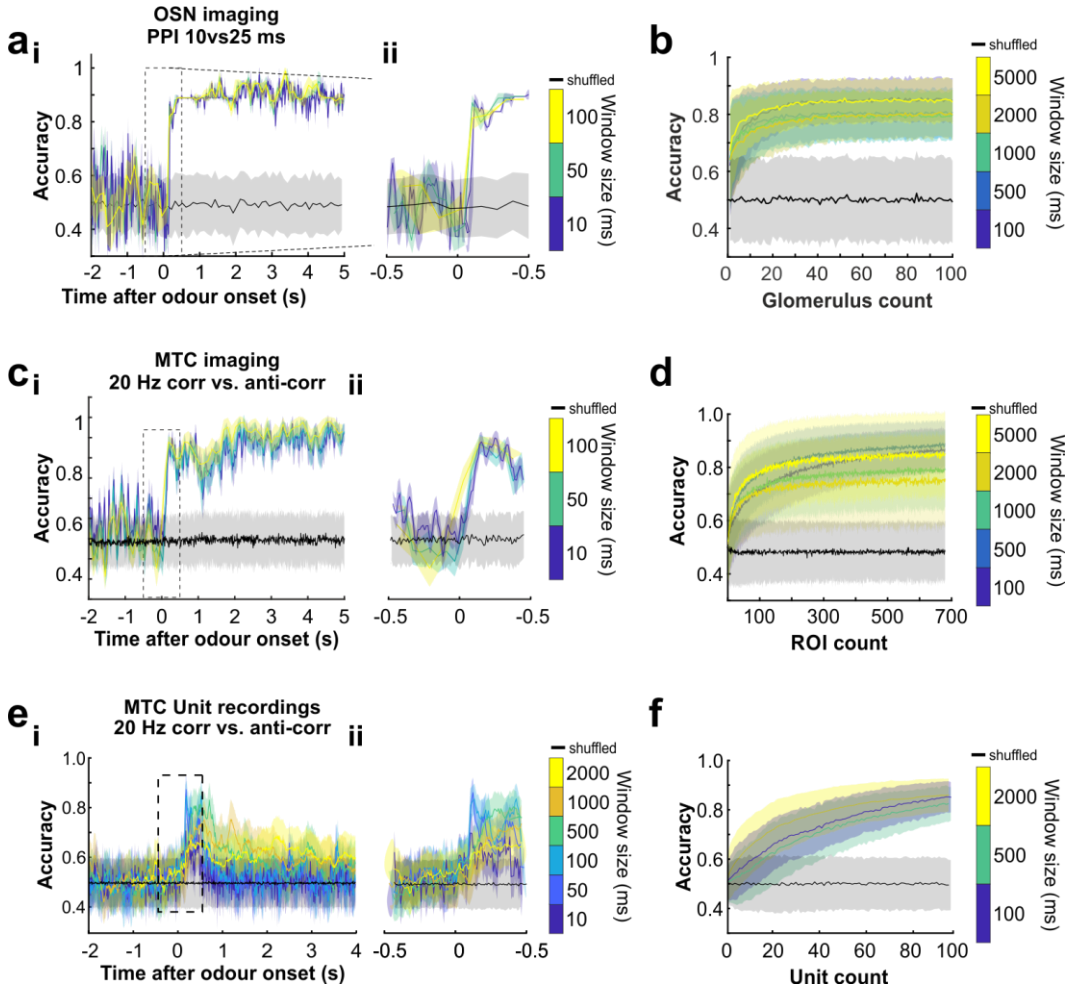

Figure R10: Classification during earlier time points of OSN imaging, and M/TC imaging and unit recordings. **a_i,** Classifier accuracy when trained on all glomeruli in response to PPI 10 vs. 25 ms stimuli recorded in awake animals (n = 100, from 5 individual animals) with a sliding window of different durations (colour-coded; black: shuffle control; mean ± SD of 100 repetitions) starting at 2 s before odour onset. **a_{ii},** Time period between

-0.5 and 0.5 s from odour onset shown at higher magnification. **b,** Classifier accuracy over an increasing number of glomeruli when trained on several response windows (colour-coded) to PPI 10 vs. 25 ms stimuli (up to 100 glomeruli from 5 individual animals; black: shuffle control; mean ± SD of 500 repetitions). **c,** Same as in **a** but when a classifier is trained on all imaged M/TCs (n = 680, from 6 individual anaesthetized animals) in response to 20 Hz correlated vs. anti-correlated stimulation. **d,** Classifier accuracy over an increasing number of M/TCs as described in **b**. **e,** Same as in **c** but when a classifier was trained on all recorded units in response to 20 Hz correlated vs. anti- correlated odour stimulation after subtracting baseline activity. **f**, Same as in **d** for unit recordings (n = 97 units from 6 individual animals).

4. The evidence for differential responses of M/T cells to odorant fluctuations is key. Unfortunately, the presentation of the data is confusing and the data analysis and discussionof the relationship to published data could be improved. First, the presentation of the data are confusing. The line for OSN decoding cannot be discerned in Fig. 3f and inclusion of bothOSN terminal and M/T cell data in Extended Fig. 8 is confusing. Extended Fig. 8 should be split into two figures (OSN input vs M/T responses) and the figure legend should be edited to provide a better explanation of the relevant information.

We sincerely apologise for the poor presentation of this point that we agree is very important indeed. Following the reviewer's advice, we have now split **Extended Data Figure 8** into three figures (**Extended Data Figure 8**: OSN imaging, **Extended Data Figure 10**: M/TC unit recordings, **Extended Data Figure 11**: Whole-cell recordings) andexpanded the description of these experiments. Furthermore, as elaborated above, we have included substantial new data for both the OSN measurements as well as for unit recordings with new high-frequency stimulus sets that in our opinion further clarifies this point.

Furthermore, the analysis of extracellular recording can be improved. The example unit in Fig. gii appears to clearly differentiate between correlated and uncorrelated.

In fact, the response of that unit resembles responses of single M/T units found to respond to 10-200 msec duration of optogenetic activation of glomerular input by Li et. al. J. Neurosci., November 26, 2014 34(48):16058 –16064. Are these data concordant with the data of Li et al?

Thank you for pointing this out – we have now included more example units in the supplement in the new **Extended Data Figure 10** to give a more objective picture of typical responses (reproduced here as **(Figure R11)**).
Also, the Li *et al.* paper is indeed one of the best examples of how temporally precise inputinto the olfactory bulb can result in distinct firing patterns – it is indeed puzzling to us atwhat point we accidentally had dropped the reference from our original manuscript. Thefact that Li *et al.* found that different M/TCs are differentially sensitive to the duration ofinput stimuli in our opinion is indeed in agreement with the fact that we do see diversityin M/TC sensitivity to differentially temporally modulated odour stimuli. We now highlight this in the revised manuscript in line 227-8 and apologise for the earlieromission.

Furthermore, the criterion for differential response of single units is not stated. We apologize for this omission. We recognize that a binary distinction of "different vs. not different" is indeed difficult to make and have included a detailed description and new analysis for the differential response of individual units whilst emphasizing that the decoding analysis includes all units we had recorded (starting in line 1593).

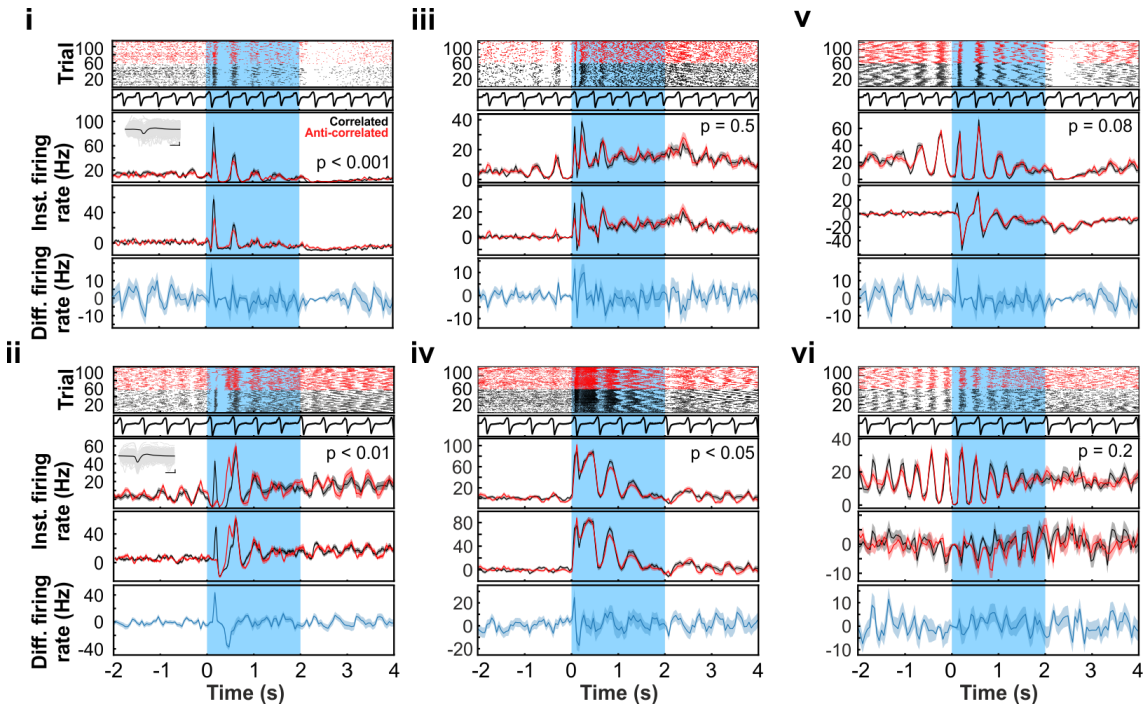

Figure R11: Single unit examples. Six additional example single units in response to 20 Hz correlated (black) and anti-correlated (red) stimuli shown as raster plots (top) and PSTH (mean of 64 trials for each condition ± SEM) of spike times before (second from top) and after baseline subtraction (second to bottom), and the differential PSTH of correlated and anti-correlated (bottom, blue). The duration of odour presentation (2 s) is indicated in light blue. PSTHs are constructed by summing the number of detected spikes in a 10 ms window multiplied by 100 to obtain the instantaneous firing rate in Hz. P-values are derived from a Mann-Whitney U test comparing the spike time distributions of correlated and anti-correlated trials during 4 s after odour onset. 24% of single units (23/97) units were classified as significantly different ($p < 0.01$).

Decoding analysis for low unit numbers is not provided.

We have now analysed a subset of units that "responded differentially" to correlated andanti-correlated stimuli as described above. We realize that this is a difficult analysis to make, as one risks to select units based on the same property one uses to subsequently analyse their population encoding. In order to avoid this potentially circular argument, we trained a classifier on a subset of repetitions in order to rank units based on their individual "response accuracy" (i.e. how different a given unit responds to correlated vs.anti-correlated stimuli, **Figure R12a**). We then derived population classifiers on the *n* "best" individual units. In order to avoid any circularity, the data shown in **Figure R12b** is the accuracy of such classifiers on fully withheld data. Indeed, even with this conservative approach, classifiers trained on "differentially responding units" outperform classifiers trained on random subsets. Further experiments will be needed to determine whether these differentially responding units might constitute a physiologically, morphologically, or molecularly identifiable subset of projection neurons. We now briefly mention this in the revised manuscript on line 264.

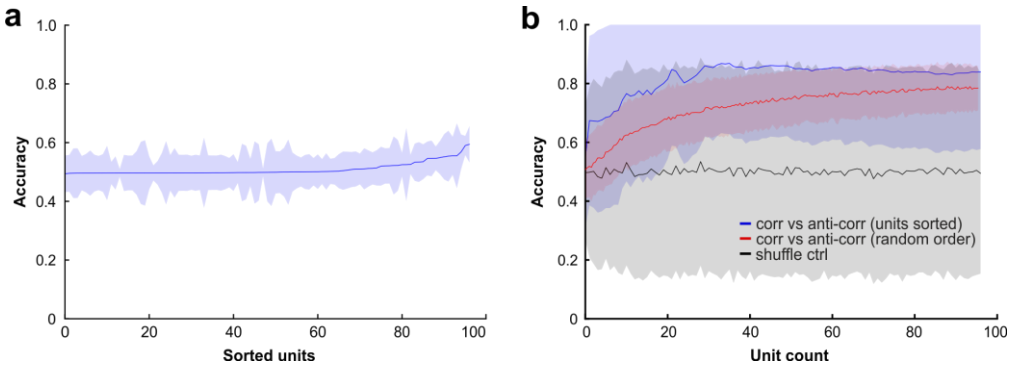

Figure R12: Units sorted by classification accuracy a, Accuracy obtained by a series of classifiers trained and tested on a single unit's response to a total of 128 trials (64 correlated and 64 anti-correlated trials; n = 97 units from 6 individual animals). Each classifier is trained on 80% of the data (102 trials) and tested on all but two of the remaining trials (24 trials). In both the training and test data, the number of correlated trials is equal to the number of anti-correlated trials (51 and 12, respectively). The training and testing split is repeated 1000 times for each unit. The units were then sorted and plotted from lowest to highest average accuracy. **b,** Series of classifiers trained and tested on the two held out trials from **a** over an increasing number of units. Units following the sorted sequence derived from **a** (blue) or a randomly selected sequence (red); shuffled labels control (black).Each classifier is repeated 1000 times with training and test data selected differently each time (mean ± SD).

Also, the PCA analysis could be useful, but as presented contributes little to the validation ofthe M/T responses. Why is the accuracy calculated through PCA analysis starkly different tothat of classifiers trained on summed spikes with small windows?

Thank you for pointing out this source of confusion. The description of the PCA approachand its motivation was indeed very compact in the original manuscript. We have now substantially expanded this in the Methods section of the revised manuscript starting line1601.
In short, training a classifier on summed spikes with very small sliding windows suffers from noise, i.e. that the spike count in e.g. a 10 ms window typically will only be 0 or 1 with spikes found in adjacent bins in repetitions. Concatenating small windows on the other hand (where in principle information from multiple 10 ms time bins would be combined without losing temporal resolution) would result in a large number of fit parameters (e.g. 100 per unit for a 10 ms time bin and 1 sec period of analysis) and consequently to an overfitting problem. Extracting key temporal features of the data through PCA allowed us to limit the number of parameters that feed into the classifier without sacrificing temporal resolution by restricting the analysis to the first *n* PCs. Thus
– in the absence of a number of repetitions that is similar to the number of time bins x units – we felt that such analysis would be most suited to assess whether increasing the resolution of analysis could be beneficial.

Finally, the whole cell data in Extended Fig. 8 appear to provide evidence for differential responses to correlated vs. anticorrelated input. Could a subset of these M/T cells beclassified as cells that respond differentially to the two stimuli?

We indeed think that this is the case. However, the difficulty of the whole-cell recordingsand the resultantly low yield makes a global statistical analysis (like we did for the imaging and unit recordings) difficult. We have, however, now assessed in more detail how individual cells respond differentially to correlated and anti-correlated stimuli. We now include data in the new **Extended Data Figure 11** that shows that indeed whole- cell recordings also reveal a subpopulation of cells that respond differentially to the two stimuli (shown here as **Figure R13)**.

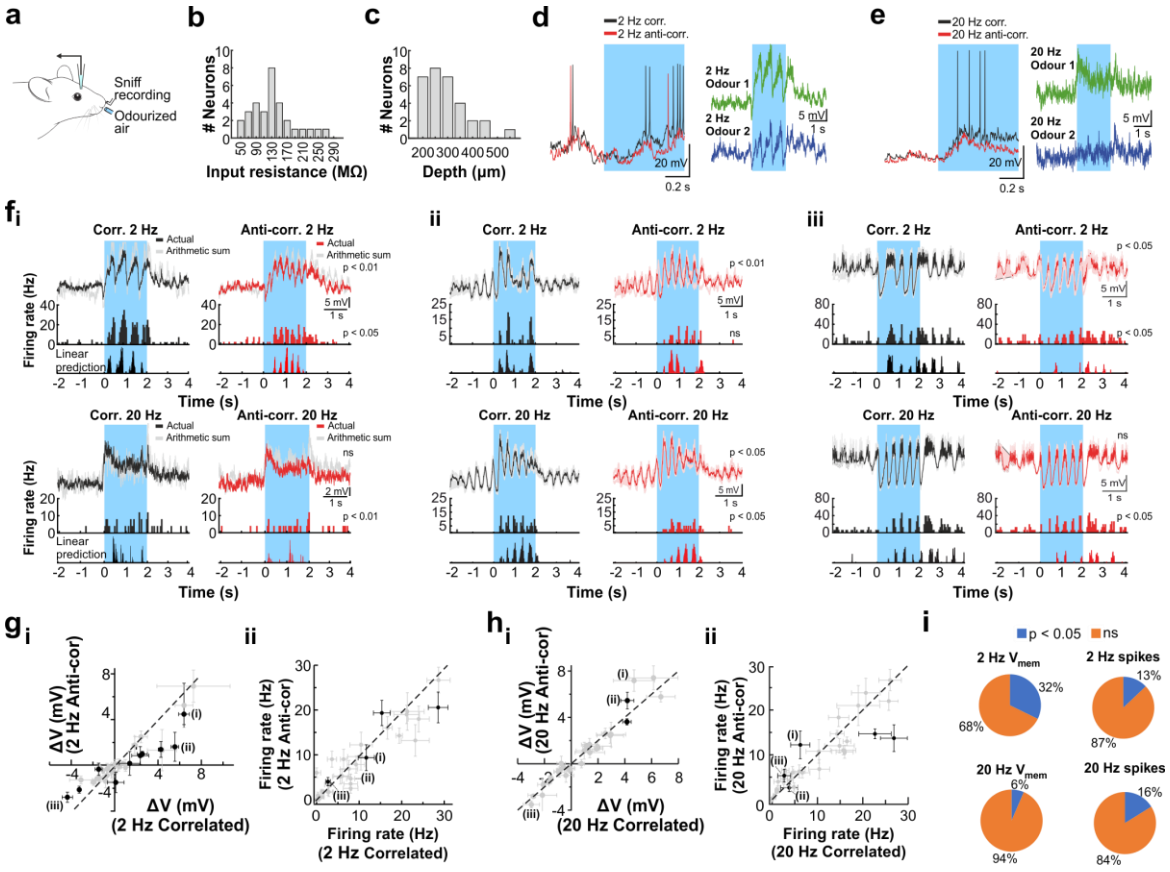

Figure R13: Whole-cell patch-clamp recordings from M/TCs in response to correlated vs. anti-correlated odour stimulation. a, Schematic of the whole-cell patch clamp recording approach. **b,** Distribution of input resistance and **c,** recording depth as measured from all recorded projection neurons (n = 31). **d,** Left: Example recording from a single cell with consecutive presentations of correlated (black) and anti-correlated (red) odour stimulus at 2 Hz. Duration of odour presentation (2 s) is indicated in light blue. Right: Baseline-subtracted and spike-clipped subthreshold voltage response from a single cell to odour 1 (green) and odour 2 (blue) for 2 Hz. **e,** the same as **d** but for 20 Hz odour stimulation in a different example cell. **f,** Voltage response from three example cells for correlated (black) and anti-correlated (red) odour stimulus for 2 Hz (top 3 rows) and 20 Hz (bottom 3 rows). The cells shown in $\mathbf{f}_i$ correspond to the cells shown in **d** and **e**. The grey overlaid traces correspond to the arithmetic sum estimated from the response to individual odours. Middle: PSTH of action potential firing. Bottom ("Linear prediction"): PSTH calculated by thresholding the arithmetic sum of the subthreshold responses to the individual odours (arbitrary units). The threshold simulates the non-linearity of action potential firing. Differences in these simulated PSTHs suggest that stimulus correlation could be calculated on a single cell level (the same inputs, phase-shifted, result in distinct simulated "action potential patterns", due to thresholding). P-values are derived from a paired t-test of the membrane potential and the mean firing rate in the first 500 ms after odour onset **g,** Average change in voltage **($\mathbf{g_i}$)** and action potential firing rate **($\mathbf{g_{ii}}$)** in the first 500 ms after odour onset from baseline membrane potential for 2 Hz correlated vs. anti-correlated odour presentation and **h,** for 20 Hz. Each marker corresponds to a single cell, error bars represent SEM. Data points in black represent cells where p

< 0.05 between the correlated and anti-correlated condition. Indicators (i), (ii) & (iii) represent cells shown in **f**. **i,** Pie chart depicting the proportion of cells showing significant difference (blue) in subthreshold membrane potential (left) and spike frequency (right) for all 2 Hz (top) and 20 Hz (bottom) cells (n = 31).

5. The section on the method for generation of odor plumes from nearby and distant pointsis succinct and does not provide enough information for the reader to understand the measurement. Please provide diagrams of the indoors and outdoors setups. Given that thereis a full speed fan drawing air into the PID will the indoors setup reproduce a situation thatan animal is likely to encounter?

This is again a very good point. We have now added detailed diagrams to the revised **Extended Data Figure 12** replicated here as **Figure R14**.

Regarding the second point, PID recordings indeed – as pointed out by the reviewer – canperturb the air flow in particular on small spatial scales. The pump rate of each PID - as we have used it - was approximately 0.02 l/s. The dimensions of the indoor as well as outdoor setting were an inter-source distance (for the separate source case) of ~50 cm. Measurements occurred at ~4 cm height. Outdoor wind speeds were ~3-5 m/s. Thus, wecan (very roughly) estimate the overall relevant air volume to be approximately 50 cm x2 x 4 cm x 4 m/s = 160 l/s. For indoor measurements the fan was set to maximum speedsuch that it pushed approximately 550 cf/min (approximately 260 l/s) which has led us to believe that the possible perturbations induced by the PID are small in these specific settings. Nevertheless, to assess the correlation structure of odour stimuli without such perturbations, we have performed new computational fluid dynamics (CFD) simulationsof two odour sources at varying distance and started to analyse the information content in different frequency bands in the absence of any PID induced perturbations to air flow **(Figure R15).**

We now give more detail on setups and PID measurements including these flow estimatesin the manuscript lines 1570-1621 and in the revised **Extended Data Figure 12**.

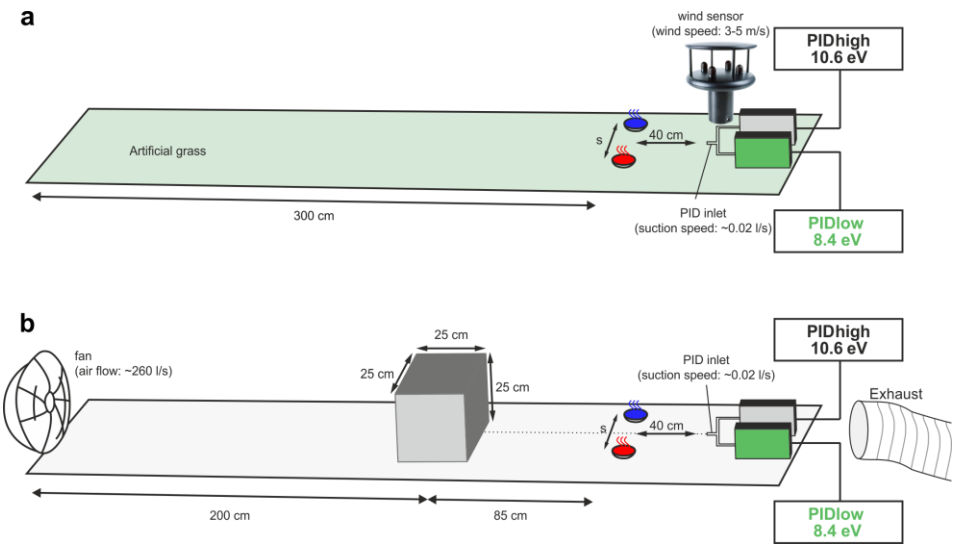

Figure R14: Schematics of outdoor and indoor plume recording setups. a, Outdoors setup: PIDs and odour delivery system were used to record multiple trials at different lateral distances (s) between odours held in ceramic crucibles. Data was collected on a day with low wind (~8-12 mph or ~3-5 m/s, recorded with a 2-axis ultrasonic wind sensor at the height of the PID inlet. Outdoor experiments were performed on a ~6 m x 10 m wooden patio structure surrounded by trees. There was >300 cm of unobstructed space on an artificial grass matin front of the PIDs to capture air movements. **b,** Indoors setup: A digitally controlled fan was placed at a distance of 325 cm facing the PID inlet. An exhaust line was situated behind the PID inlet to ensure the direction of air from the fan towards the PID inlet. During a recording, the fan was set to maximum speed such that it pushed approximately 550 cf/min (cubic feet per minute, ~260 l/s) of air towards the PID inlet. A 25x25x25 cm Thermocool box was placed 200 cm downwind of the fan acting as an obstacle to air movement, breaking up any laminar airflow and ensuring complex, turbulent-like air movement at the PID inlet. The pump at each PID was set to ~0.02 l/s suction speed, unlikely to perturb overall airflow dynamics substantially.

Finally, please do not state that these conditions are "turbulent" unless this statement is backed by the physics underlying turbulent flow. "Complex" flow patterns may be more suitable.

This was indeed imprecise language and we apologize. We can indeed not be sure that conditions were turbulent as with the point measurements we have made with the PIDs we cannot check many of the key characteristics (vortex structure, diffusivity, dissipation…). We thus follow the reviewers advise and have – as suggested – reverted tothe term "complex" (e.g. lines 25, 48, 186, 277, 442, 449, 948).
Furthermore, the CFD simulations mentioned above have allowed us to assess air flow ona larger spatial scale (compared to the point measurements of the PIDs). While these simulations might similarly not produce properly turbulent airflow (as they are 2- dimensional), they do display vortices of different scale and result in temporal patterns of odours carried within these complex airstreams similar to those observed experimentally **(Figure R15a,b)**. We used these simulated flow patterns and assessed the structure of odour fluctuations when multiple odour sources were present. As experimentally observed, correlations between odour fluctuations decreased with increasing inter-source distance **(Figure R15c)**. While in ongoing work we now quantifyinformation content of different temporal features, for the purpose of this manuscript wesolely see these controlled simulations as evidence that complex airflow patterns result in high-frequency odour fluctuation patterns that contain information about odour sources in the scenery.

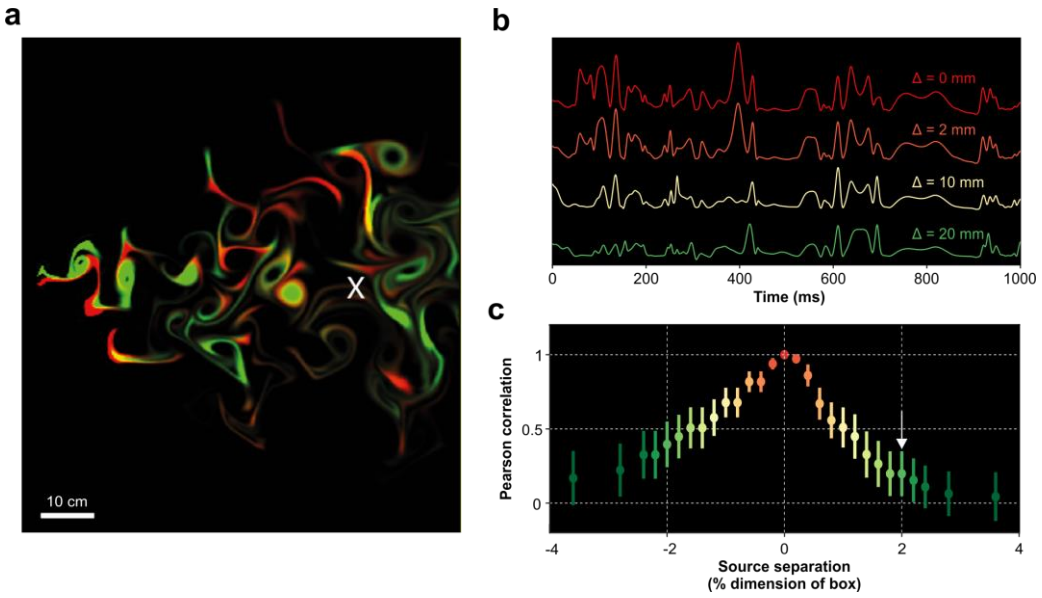

Figure R15: Computational fluid dynamics (CFD) simulation of odour plumes. a, Example 2D CFD simulation showing plumes produced by airflow flowing at 1 m/s past obstacles (not shown), then over two point odour sources (red, green points centre left) placed 20 mm apart. Obstacles are 12 cylinders of 38 mm diameter equally spaced vertically with a gap of 38 mm between neighbouring cylinders, horizontally centred 20 cm to the left of the odour sources. 'X' marks the location of the concentration probe used in the remaining panels and is 60 cm directly downwind of the central (red) source. **b,** Snapshot of plume concentration profiles for sources of increasing displacement. Most top and bottom traces correspond to the red and green plumes shown in **a**, respectively. **c,** Correlation between concentration profiles from two sources as a function of source separation. Mean (points) ± SD (lines) of the Pearson correlation between the concentration profile of the central source (red in **a**) and sources vertically displaced from it, measured at the location 'X'. Correlations were computed in sliding1 second windows with an overlap of 500 ms between adjacent windows, for the last 8 seconds of simulation. The arrow indicates the correlation between the signals from the red and green plumes in **a**.

Finally, is the outdoors correlation significantly different from one for the one source case?Did this differ for different volatiles?

This is again an important point that we didn't make clear in the original manuscript. Firstly, yes, the correlation is significantly higher for one source vs. 50 cm separation (0.67 ± 0.11 vs. 0.24 ± 0.24, mean ± SD, $p < 0.001$, Two-sample t-test, **Figure 4f**). We apologize for not having highlighted this important point and now clarify this in the figureand the corresponding legend in the revised manuscript.
However, the differences we measure in the outdoor setting will significantly underestimate the actual differences in correlation of the two odours in question: In the outdoor setting there will be background odours, both (i) with ionisation energy below the "PIDlow" ionisation energy of 8.4 eV and (ii) with ionisation energy between "PIDlow"(8.4 eV) and "PIDhigh" (10.6 eV). (i) will result in a (identical except for a scaling factor) signal in both PIDs as both the 10.6 eV and the 8.4 eV bulb will ionize these odours, whereas (ii) will result in a signal only in "PIDhigh" (as 8.4 eV will not be enough to ionizeit). Thus, background odour (i) will increase the apparent odour correlation whereas (ii)will decrease it. Therefore, even if odour A and B were perfectly correlated in their fluctuations for the "one source" case, the correlation measured would be below 1 due tothe presence of background odour (ii). Similarly, even if A and B were fully uncorrelated,background odour (i) would result in residual apparent correlation. Thus, the difference between the outdoor measurement and the indoor measurement seems consistent with the presence of background odours.
We would not expect this to be different for other volatiles, however there is a limited setof odour pairs we can use with the dual-PID approach we have developed (limited by the requirement for sufficiently different ionisation energies and the fact that the reagents that need to freely spread in air should not be toxic to equipment or experimenter). This,and the fact that the lower ionisation energy UV bulbs are less powerful, thus resulting inworse signal to noise meant that we decided to experimentally test different pairs of volatiles in the high signal-to-noise indoor laboratory environment only.
We now detail this reasoning in lines 1501-1517 of the revised manuscript.

6. A substantial number of statements are not backed up by statistical analysis. For examplethe claim that mice perform above chance for probe trials in Figs. 4j,k is not backed up by astatistical test. For all figures where there is a comparison please show statistical differenceswith an asterisk and in the legend provide all the information for the statistical test used. For example, for Fig. 4di are some correlations statistically different? Which multivatiate test was used? What was the sample size? Are the data distributed normally? Are interquartiles shown (all that is stated in the legend is that what is shown is mean+SD)? Please revise the statistics through the document.
We sincerely apologise for these omissions. We have now included statistical descriptions in much more detail overall and in particular in the **Figures 1f,k, 2c,p, 3d,e,m, 4d,e,f,j,k**.

Minor comments

1. Line 44 change "prowess" to "powers"
Thank you – we have changed to "resources"

2. Line 54, this reference should be to Fig. 1c
Thanks for spotting this - done

3. Line 133, please break into two paragraphs. Please add blank line between paragraphs.
done

4. The example of the ROI in Fig. 3c appears to be a dendrite. Yet, in the methods it is statedthat all measurements were from somata. In fact, the information content may differ between somata and the dendrites within the glomerulus. Did the authors perform evaluation of dendritic responses for the M/T Cells?
Thanks for raising this interesting point. The ROI depicted in the figure is a soma – although this is indeed impossible to see from the figure. We have now included a betterlabelled **Figure 3c** with magnified insets in the revised manuscript.
However, the reviewer question prompted us to specifically analyse dendrites as well. Indeed, dendrites can readily be segmented from our 2p stacks **(Figure R16a)**. Dendriticsegments responded reliably to the correlated and anti-correlated stimuli **(Figure R16b,c).** We then performed SVM classifier analysis as for the data extracted from somata. Interestingly, classifier accuracy in fact exceeded that obtained from soma measurements although it reached these performance levels more slowly **(Figure R16d,e)**.
We now show this data in the new **Extended Data Figure 9** and comment on it in line 148 of the revised manuscript.

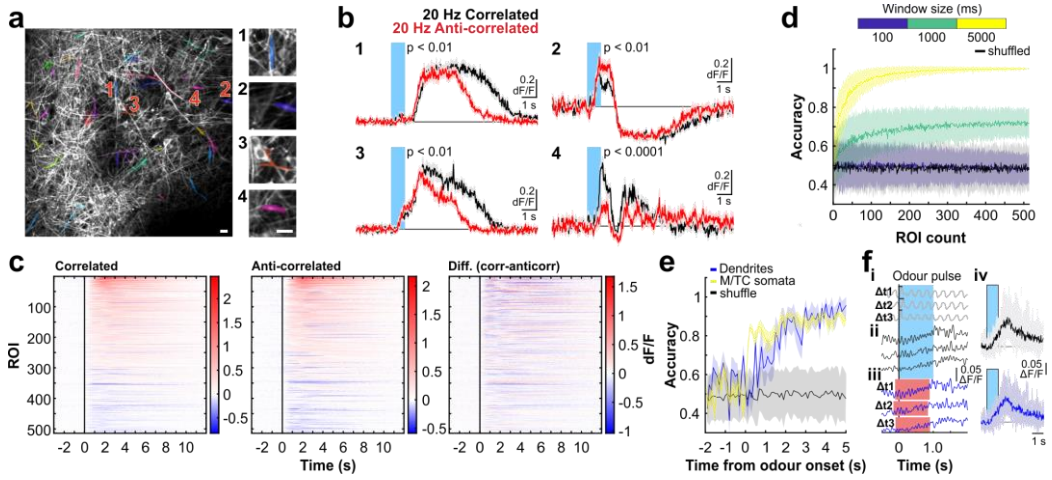

Figure R16: Odour correlation structure is encoded in dendrites of olfactory bulb output neurons. a, GCaMP6f fluorescence from mitral and tufted cells and their dendrites recorded in the dorsal portion of the olfactory bulb of a Tbet-cre:Rosa-GCaMP6f mouse (maximum projection of 8000 frames). Dendritic ROIs are superimposed in colour. Four dendritic segments (1-4) are shown at higher magnification, scale bars: 20 µm. **b,** Four example calcium traces extracted from dendritic segments shown in **a** that show differential response kinetics to correlated (black) and anti-correlated (red) stimulation (mean of 24 trials ± SEM, f = 20 Hz). In total, 23.54% of all dendritic segments analysed showed significantly different integral responses to the two stimuli (0-5 s after odour onset, $p < 0.01$, unpaired t-test; 121/514). **c,** Calcium transients as colour maps for correlated (left) anti-correlated (middle) and the difference between both odour stimulations (right) of all analysed dendritic segments (n = 514, from 6 individual animals). **d,** Classifier accuracy over an increasing number of dendritic ROIs when a linear classifier was trained on several response windows (colour-coded) to correlated vs. anti-correlated stimuli at 20 Hz (n = 514, mean ± SD from 6 individual animals, black: shuffle control). **e,** Classifier accuracy when trained on all dendritic ROIs recorded with a 100 ms sliding window starting 2 seconds before odour onset (mean

± SD, 100 repetitions), classifier for somatic M/TC ROIs superimposed in yellow (from Figure 3g, 100 ms sliding window). **f,** Method of aligning calcium traces to first inhalation after odour stimulus onset. **f_i,** Representative respiration traces recorded using a flow sensor placed in front of the nostril contralateral to the imaging window. The first inhalation peaks were detected and the time (Δt) to the first inhalation after odour onset was calculatedfor each trial individually. **f_{ii},** Representative calcium transients in response to a single odour presentation (here:20 Hz correlated). **f_{iii},** Transients are shifted according to Δt. **f_{iv},** Individual calcium transients (faint colours, 24 trials) in response to 20 Hz correlated odour presentations with the average calcium signal (thick traces) superimposed. Top: before aligning to first inhalation after odour onset, bottom: after alignment. Blue bar represents the 1 s odour presentation phase (approximate for the aligned data).

5. Line 155: "Consistent with the Ca2+ imaging results, single-units did not follow high- frequency stimuli, but the average firing rate of 19% of units (18/97) again differedbetween correlated and anti-correlated stimuli (Fig. 3m,n and Extended Data Fig. 8g,h)." Please modify the sentence. First: Ca2+ imaging does not follow high-frequency stimuli because of technical issues. Second: What is the basis for classifying these as units that display a statistically significant difference in firing.
Thank you. We have now corrected the sentence and added more detail on the classification in the revised manuscript in lines 159-164.

6. Line 393: "o, Average accuracies for sets of linear classifiers trained with different rolling window sizes (colour-coded), when the number of units available for the classifier was varied (n = 97, mean +/- SEM from 6 individual animals; see also Methods and ExtendedData Fig. 8)." This figure shows the dependence of accuracy on window size as opposed tothe number of units. Please show separate graphs for window size and number of units. Thank you for pointing out this source of confusion. We now plot accuracy as a functionof the number of units in the revised **Figure 3**. In the main figure we do this for threedifferent time windows and provide a more detailed analysis in the new **Extended DataFigure 10**.

7. It is surprising that for the awake recordings in Extended Data Fig. 8 b the animal delays breathing for nearly one second. This reminds me of the response to irritant stimuli. Could it be that the concentration of the stimuli is so high that it is being detected as an irritant? Did the awake mice slow their breathing when presented with the odorants?
Thank you for spotting this. We have now reanalysed the sniff behaviour in the recordings **(Figure R17)**. It turns out that the apparent pause in breathing shown in the example inthe former Extended Data Figure 8b was a very unusual and rare case. As expected from the literature, sniff rate in fact increases during odour presentation **(Figure R17ai-iii)**. We have, therefore, now chosen more typical examples in the new **Extended Data Figure**
8. We have also included a detailed sniff analysis of our awake and anaesthetized recordings as the new **Extended Data Figure 3h**.

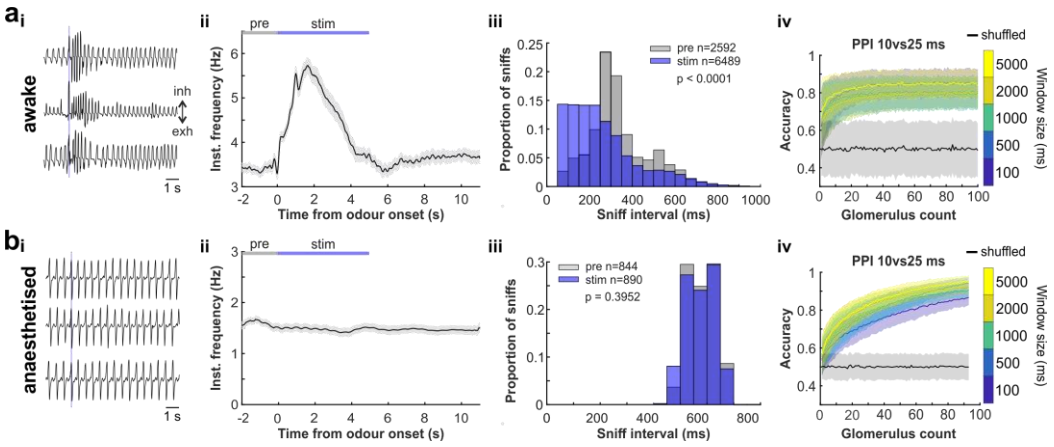

Figure R17: Respiration pattern analysis in awake and anaesthetised mice. a_i, Example respiration traces recorded from an awake mouse using a flow sensor. Inhalation goes in the upwards, exhalation in the downwards direction. **a_{ii},** Average instantaneous sniff frequency from one example animal plotted as a function of time (n = 24 trials, mean ± SEM). The odour stimulus consisted of two 10 ms long odour pulses either 10 or 25 ms apart (see Figure 1c). **a_{iii},** Distribution of sniff intervals during a 2 second window before (grey) and a 5 s window after(blue) odour stimulus onset ($p < 0.0001$, two-sample Kolmogorov-Smirnov test). **a_{iv},** Classifier accuracy over all glomeruli when a linear classifier was trained on several response windows (colour-coded) to PPI 10 vs. 25 ms stimuli (n = 100, from 5 individual animals; black: shuffle control). **b,** same for the anaesthetised condition (**b_{iii},** p

= 0.3952, two-sample Kolmogorov-Smirnov test; **b_{iv},** n = 93 glomeruli from 4 individual animals).

8. Please state how many animals were tested for each measurement.
Done – we have now included the number of animals in each figure legend.

9. Line 141. How was the significant difference determined for the integral responses? It seems integrating the response would loose temporal information. Were the responses subtracted before calculating the delta response integral?
Neural responses between correlated and anti-correlated trials were determined to be significantly different if response integrals calculated over a set of windows after odour onset showed p-values < 0.01 (unpaired t-test). This was the case for 17% of all M/TCs (5000 ms window: 114/680). In addition, in order to keep more temporal information, we have now analysed the response integrated also over shorter time windows: 100 ms (6%), 500 ms (8%), 1000 ms (9%) and 2000 ms (12%). We now highlight this transparently in the results section of the revised manuscript, line 145.

10. Fig. 3n provides no information on differential responses. Please provide a raster of the delta PSTH.
We have now included a raster of delta PSTH in the revised **Figure 3m** as well as a heatmap highlighting the differential responses across all recorded units in **Figure 3n**.

11. The time scale bar is missing in Fig. 4g.
Thank you – now included in the revised **Figure 4g**.

12. Line 198. On what basis is this statement made: "Notably, they reliably categorized correlated stimuli as "same source" and uncorrelated as "source separated""
After training mice on "same source" and "source separated" stimuli we probe their performance with e.g. correlated pulse trains or recreated "natural plume dynamics" odour stimuli (with two odours perfectly correlated). Animals respond to those probe trials as they did for the "same source" trials before – we have rephrased the passage in question (line 206 in the revised manuscript) now to make this explicit.

13. Line 197. "that were derived from turbulence measurements". Were these really measurements of turbulence?
Thanks for pointing out this poor choice of words. We have now rephrased this sentencein the revised manuscript (line 186) and – as discussed above – replaced turbulence withthe more accurate description "complex airflow"

Referee #3 (Remarks to the Author):

Schaefer and colleagues tackle an exciting, important and underexplored issue in sensory neuroscience, specifically olfaction. It is often not appreciated that chemical stimuli in natural conditions arrive at sensors in fluctuating, intermittent manner due to the inevitable physics of the world. These fluctuations have a lot of high frequency componentsthat can be dismissed as noise that needs to be averaged out. However, could animals makeuse of these fluctuations to infer important things about the olfactory world? In this paper,the authors offer evidence that mice can make use of high frequency olfactory information to make decisions. Related, the study also offers a demonstration of the idea that decorrelation across time can be exploited behaviorally (by mice). In addition, the authors also offer some support for neural coding of high frequency information in olfactory stimuli.

Overall, this is an intriguing study that is both creative and novel. The behavioral work is rigorous and convincing, but I didn't find the neural recordings all that convincing (but it doesn't at all detract from the importance of the behavior – so not a problem!). I elaborate below with specific comments.

Thank you for the encouraging words and constructive comments! We have addressed all comments with new analysis, new experiments (in particular more and more detailed physiological experiments), improved data presentation and edits to the text as suggested.

Abstract, line 6: It seems like "as little as" should be followed by a single number, rather than10 or 25 ms.

Thank you – we have rephrased accordingly.

Line 52/54. This refers to Fig 1b

Thank you – corrected.

Line 61-67 / Extended data Fig2: The authors show that a simple biophysical model is sufficient to have neurons exhibiting different responses to the two types of pulses despitethe lack of faithful stimulus-following in time. For completeness and to be even moreconvincing they should add a simple decoding analysis similar to the one performed for OSNresponses (fig 1i,j) showing decoding as a function of network size. I also wonder ifadaptation could play some role, but perhaps worrying about all the details is not helpful! Following the reviewer's comments, we have now extended our OSN model. Performingthe same decoding analysis as for the OSN imaging data indeed shows a very similaraccuracy profile **(Figure R18g,h)**. We have also added additional biophysics-inspiredparameters, such as the olfactory transduction time constant, OSN membrane timeconstant, spike threshold and refractory period **(Figure R18g)** that can indeed aid encoding of temporal structure. We have included this new data in **Extended DataFigure 2** and discuss both of these aspects in the figure legend. We are convinced thatfurther biophysical complexity and heterogeneity such as derived from the differentfeedback and amplification stages (e.g. Ca^{2+} gated Cl^- channels) could further enableefficient encoding of different temporal structures in OSNs (similar to what had beendescribed e.g. for lobster OSNs Park et al., 2014, 2016 PMID: 24431452, 26730727). Weagree that - in particular with more OSN biophysical data - this will be an important fieldfor future studies of mammalian olfaction, and comment on this accordingly in thediscussion on line 260 of the revised manuscript.

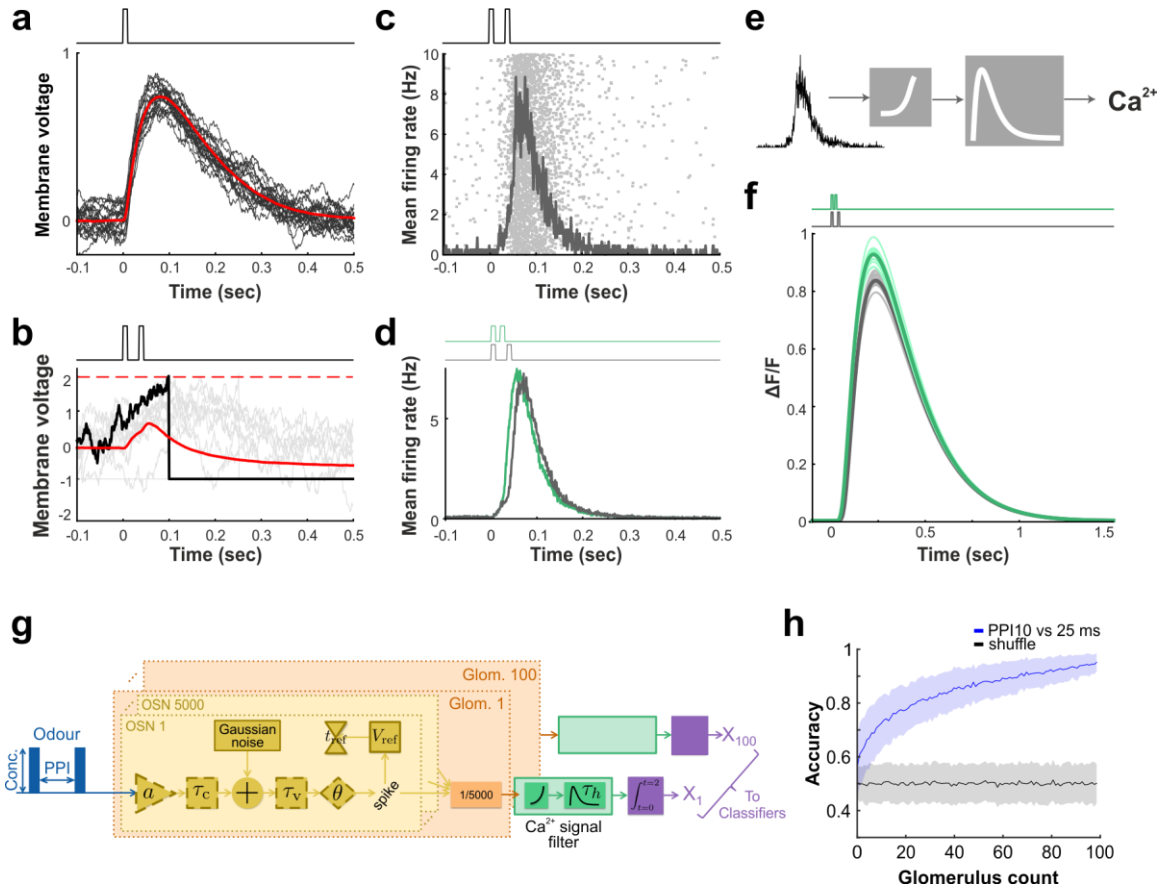

Figure R18: Distinguishing fast odour stimuli with slow OSNs. a, Membrane voltage relative to baseline of a single model OSN in response to a 10 ms odour pulse. Black traces are individual trials; red trace is the average over 20 trials. OSN spike threshold has been set high enough to prevent spiking to illustrate the subthreshold voltage time course. **b**, Membrane voltages (grey traces) of ten OSNs from a population of 5000 in response to a paired odour pulse with pulse width 10 ms and PPI (paired pulse interval) of 25 ms. Several OSNs reach the OSN spike threshold (dashed red line) and are temporarily reset to the refractory voltage of -1. The population average membrane voltage (black) reveals membrane charging in response to odour stimulation and the subsequent discharging and refractory period. **c,** Raster showing the spike times (dots) of the full population from **b** and the corresponding mean firing rate (trace). **d,** Mean firing rates computed over 20 trials in response to paired odour pulses of width 10 ms and PPIs of 10 ms (green) and 25 ms (black). **e,** Model calcium signals are produced by squaring the instantaneous mean firing rate and filtering the result with a calcium imaging kernel. **f,** Model calcium responses to the paired odour stimulus with a PPI of 10 ms (green) and 25 ms (black). Thin traces are single trials, thick traces are averages over 15 trials. **g,** Schematic of the OSN model. Variables in dashed bounding boxes are changed for each glomerulus (see Methods). **h,** Linear classifier analysis over an increasing subset size of glomeruli (1-100; plotted is mean ± SD, 256 repeats for random subsets of *n* glomeruli generating 256 unshuffled and 256 shuffled accuracies). The full code and data used to create the figure above are available at: https://github.com/stootoon/crick-osn-model-release.

Line 66: the logic here ignores concentration fluctuations - that is small differences of responses can also be interpreted as different concentrations. How will the brain know which is the cause of the different amplitude responses?

Thank you very much for pointing out this omission in our modelling results (as for the experimental data below [comment on line 81]). Indeed, based on the amplitude or integral of a *single* OSN population response, timing and concentration cannot be separated. However, *how* a response depends on concentration and timing (e.g. the shapes of the concentration-response and timing-response curves) will vary between OSN populations, such that the two parameters can be readily disambiguated.

We now show this explicitly for our modelling results **(Figure R19a,b)** and also for the experimental data from OSN recordings **(Figure R19c,d)**. We now include this in the revised **Extended Data Figure 3j,k**.

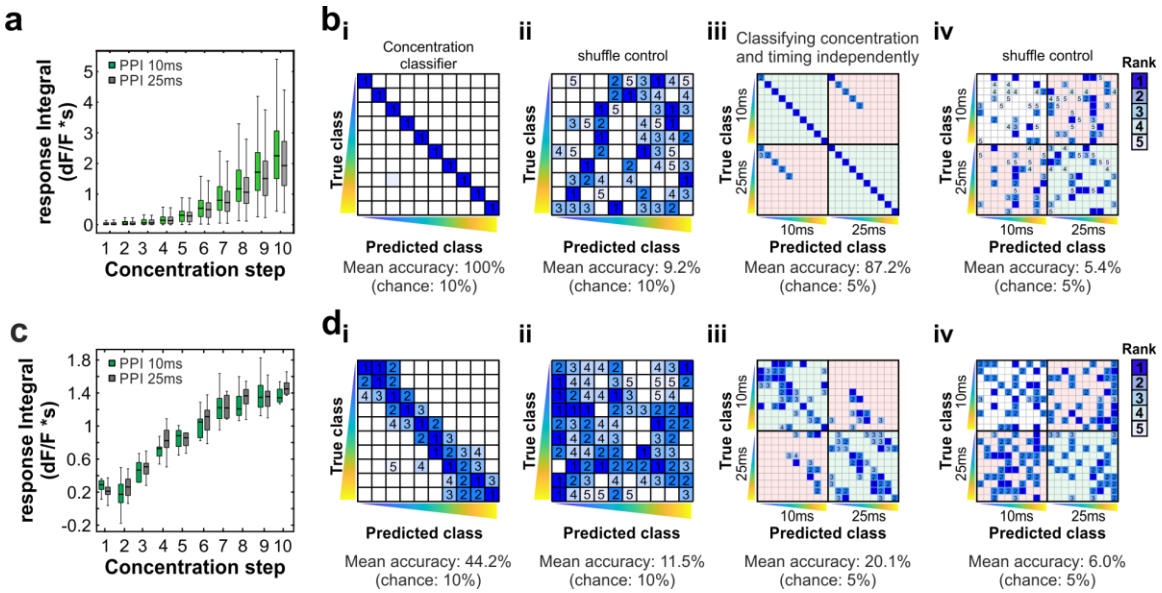

Figure R19: Classifying paired pulse interval over concentration range. a, Modelled response integrals to PPI 10 vs. 25 ms stimulations over a concentration range (10 steps) of an example glomerulus (see Extended Data Figure 3 / Figure R18 and Methods). Box plots show median of 10 repeats and 25th to 75th percentile. **b_i,** Confusion matrix of support vector machine (SVM)-based classification results of modelled glomerular signals in response to a range of 10 odour concentrations ranked and colour-coded (n = 100 glomeruli). **b_{ii},** Shuffle control with labels assigned randomly. **b_{iii},** Confusion matrix showing the ranked and colour-coded results of glomerular responses independently classified for 10 ms vs. 25 ms PPI and across the range of 10 odour concentrations. **b_{iv},** Shuffle control for **b_{iii}** with labels assigned randomly. **c,** Same as **a** but 2 s response integrals are derived from Ca^{2+} imaging data (10 repeats for each concentration). **d,** Same as **b** for Ca^{2+} imaging data (n = 57 glomeruli, from 2 individual animals, 10 repeats for each concentration). Note that 10 ms PPI could be reliably distinguished from 25 ms PPI with only few instances where a response to e.g. a 10 ms PPI stimulus was misclassified as 25 ms or vice versa (compare light red quadrants to light green quadrants).

Code and data for the model used to produce panels **a** and **b** are available at https://github.com/stootoon/crick- osn-model-release.

Line 77: I was unable to find in the methods whether classification is on single trials (and not trial-averaged). I may have overlooked this, but please clarify.

Classification was indeed on single trials – and we now point that out in the methods on line 1491/2 of the revised manuscript.

Line 81: As far as I can understand, timing and concentration are not really independently coded. If you were given two traces that are slightly different, you couldn't tell if the difference is due to timing or concentration of odor. The confusion matrix should be across modality of decoding.

This is a very good point and similar to the one raised above for the simulation data (comment on line 66) where we have answered it in detail. We have now included a confusion matrix across modality as suggested in the revised **Extended Data Figure 3k and m**.

Line 84 or so: The authors should state in the main text that the task is a go/no go design (line 84). They use as accuracy the mean of hits and correct rejects. Due to the asymmetry of the go/no-go paradigm, it would be best to show a version a plot of Fig 2k and Fig 4k separately for S+ and S- in the supplement. This would show that the increases in accuracy is properly distributed across the two options. (for example, mice could be mostly licking more or mostly licking less, with a very biased response hence the non-saturating accuracy)

We have now included the go/no go design explicitly in the revised manuscript (line 84).We have also included a version of Figure 2k and Figure 4k with separate performance on S+ and S- trials as a new **Extended Data Figure 5l and 13l** (shown here as **Figure R20**). As expected, most mistakes are made on the S- trials with the "default" response ofmotivated animals being a "Hit" (for S+) and "False Alarm" (for S-)

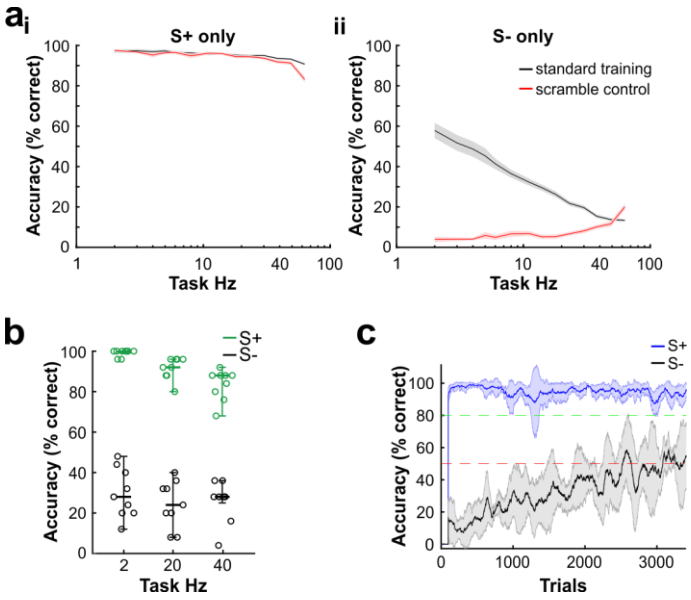

Figure R20: Correlation discrimination accuracy split up by stimulus valence a, Discrimination accuracy (mean ± SEM) for **a_i,** rewarded S+ and **a_{ii},** unrewarded S- trials with stimulus pulse frequency randomised from trial to trial when odours were presented using standard training valve configurations (black) and scrambled valve identity (red), data from Fig. 2k (mean ± SEM). Note that frequencies >40 Hz were presented predominantly in the last block of the training schedule (after several months of training) and reduced licking in the control group (decreased S+ performance and increased S- performance) might be due to decreased motivation at that point. **b,**Discrimination accuracy split by stimulus valence (green, S+; black, S-) for odour correlation fluctuation frequencies 2, 20 and 40 Hz. Data collected from the animal cohort shown in Fig. 4j. **c,** Learning curve split up by stimulus valence for an example animal trained to perform the virtual source separation task using odour plumes (see Fig. 4g). Discrimination accuracy is calculated over a 100 trial sliding window and plotted as mean ± SEM.

Line 89, figure 2c: I really worry that this result (that mice simply cannot detect flickering stimuli very well) is telling us something. The authors argue that this is a poor test of temporal abilities and go on with their correlated-anticorrelated stimuli for the rest of the paper. I applaud the rigor and the various controls the authors have done – an excellent scientific process. But it is indeed puzzling why mice can't separate out, for example, 6 Hz vs20Hz stimuli, very well. I wonder if the authors can discuss this further, along these lines (forexample):

If you correlate two traces with the same phase but different frequencies (as in stimuli usedfor Fig 2c), you'll get an envelope rather than a zero-mean correlation – this could make it harder for mice to detect. Maybe the authors could show the temporal correlation for thosestimuli? The stimuli the authors are using later (fig 2d, for example) for the correlated- anticorrelated expts are at one end of the spectrum (if you sum the stimuli, ignoring identity,the pulses go from 0 to 2 for correlated and stable at 1 for anti-correlated – could mice be using some summed intensity metric?). If you do the same calculation for the 6hz-20hz, theenvelope you get is somewhere in between – the authors could show those, and that might partly explain why mice are worse in Fig 2c?

Thank you for pointing this out – and we sincerely apologise for the confusion we have caused. For this part of the study, animals were only trained on *one* odour and only *one* frequency was presented at any given trial. I.e., the flickering stimuli were presented individually (a given trial had odour A at e.g. 10 Hz (S+), the next trial odour A at 20 Hz (S-)) not simultaneously. So, we think that the explanation is likely the same as for the visual system, namely that absolute time scales (whether a fluctuation is at 10 Hz or at 20Hz) are significantly harder to distinguish than simultaneity (two stimuli co-occurring / correlated).
We have now rephrased the respective paragraphs (line 89 in the revised manuscript). We have also dug deeper in the literature for the visual system and included more specific discussion of this aspect in lines 94-7 of the revised manuscript.

Lines 145-147: It would be good to show that non-linear classifiers, or some metric of information, can pull out the differences from glomerular data. M/T cells are not pulling outinformation out of nothing!

Thank you for raising this point - that was similarly mentioned by Reviewer 1. We duplicate our answer here for easier readability.
We have now performed more glomerular imaging experiments to get a richer dataset toanalyse and assess what information is available in the OSN Ca^{2+} signal. We recorded responses to 3 different odour pairs at frequencies of 2 Hz as well as 20 Hz. With now a total of 435 glomeruli-odour pairs (3 odour pairs, 145 glomeruli recorded across 5 animals) we find – using the same classifier analysis as before – that 2 Hz correlated vs. anti-correlated stimuli can indeed be classified with up to 70% accuracy **(Figure R1a)**. For 20 Hz stimuli classifier accuracy of ≈57% can be reached with 100+ glomeruli **(Figure R1c)**. Following the reviewer's advice, we then went on to expand on the classifiers we had used on the new data (with 3 odour pairs and more repeats). In order to reduce the number of fit parameters and therefore increase robustness to overfitting we first tried an improved linear classifier and imposed a sparseness constraint. Indeed,classifier accuracy was somewhat improved for the 2 Hz case (shown for one example odour pair in **Figure R1bi** and for all 3 odour pairs in direct comparison to the linear classifier without sparsity constraint in an appendix **Figure R21a,b**). Note that indeed – as expected if overfitting is not an issue – the shuffle control classification was robustly at chance **(Figure R1bii)**. Testing the same sparse linear classifier at 20 Hz revealed accuracy that substantially exceeded chance for individual time bins (**Figure R1ci** whilstshuffle controls remained at chance **Figure R1cii**). Classification, however, wasn't robustacross time windows or time points or odours (see also **Figure R21a** for all odour pairs).Thus, linear classifiers can indeed robustly reveal information about correlation structure at low frequencies and some information about correlation structure even at 20 Hz. This, however, is substantially less pronounced and robust than in the M/TC responses (**Figure 3** and **Extended Data Figure 9,10,11).**
Following the reviewer's suggestion, we then explored non-linear classifiers (summarized in the appendix **Figure R21c,d**). In particular, we focussed on non-linear classifiers that can be parameterized with small numbers of parameters (albeit still substantially larger than for the linear case) to reduce the risk of overfitting. Unfortunately, neither 3rd order polynomial nor radial basis function kernels substantially improved classification accuracy. In fact, both approaches consistently resulted in below-chance classification accuracy on withheld training data for the 20 Hz case as well as for all shuffle controls **(Figure R21cd)**, indicating potential overfitting. So,in short, while non-linear classifiers might be able to pull out information about stimulus

correlations more effectively, the large number of parameters make training such classifiers very difficult. However, already with simply more data and more in-depth linear analyses we could indeed pull out information about correlation structure from glomerular imaging. Moreover, this analysis has triggered us to more closely investigate responses of individual example glomeruli that indeed show that at least some information about odour correlation can be extracted from OSN imaging **(Figure R5)**. We should emphasize, however, that all this analysis is on slow Ca^{2+} imaging data, averaged across all OSN axons in a given glomerulus. The neural circuitry of the olfactory bulb of course has access to much more fine-grained information from OSN activity – action potentials with millisecond precision and thousands of individual fibres. We havenow included this new data and new analysis in the new **Extended Data Figure 8** in therevised manuscript.

Figure 3: All the physiology seems to be done in rhythmically breathing animals, where things are more predictable. As seen in sniffing traces during behavior, there is natural fluctuation in breathing that may make things harder to decipher.

We have now performed more detailed analysis of sniff variability in our head-fixed awake physiology experiments that is similar to what has been reported elsewhere for awake mice (e.g. Wesson *et al.,* 2008, PMID 18399719; Roland *et al.,* 2016, PMID 27177421; **Figure R17**) – indeed some of the examples in the original manuscriptmisleadingly suggested continuously rhythmic breathing which is not the case across thepopulation of awake recordings.
The reviewer rightly points out that such variability in sniff duration indeed has often made averaging of traces, decoding, and analysis in general more difficult. Therefore, forrecordings in awake animals, the odour stimulus was triggered on inhalation onset. In theshort odour pulse experiments (**Figure 1** and **Extended Data Figures 3,10**), odour waspresented within a single inhalation (see **Figure 1c**), greatly reducing response variability compared to stimulation durations across multiple sniffs. Moreover, directly comparing awake and anaesthetized recordings (e.g. **Figure R17**) shows that there are no qualitative differences in decoding.
We now show the new sniff data analysis in a new panel of the revised **Extended Data Figure 3h** and the sniff-locked stimulus presentation and alignment strategy in **Figure 1** and in the new **Extended Data Figure 9f**.

Line 169: I'm not sure if one can actually ATTRIBUTE the location (that happens when the agent gets closer). It's more that one can say they are separate sources (or the same source,spatially).

Thank you for pointing out this inaccuracy – we have now rephrased more precisely "attributing the various chemicals present to the same or different objects" in line 174 ofthe revised manuscript.

Line 191: once again - the information is NOT about the location, it's more about single or multiple sources

Thank you again – we have now rephrased more precisely "…contains reliable information about the composition of odour objects – whether odours emerge from the same or different sources." in line 197 of the revised manuscript.

Line 195: It looks like mice take a long time to learn... does this mean that this is actually hard to do, and not very natural for mice?

We think that the fact that it takes animals several days to acquire significant performance is largely due to learning operant aspects of the task and the "low-pressure"home cage learning environment (self-paced, no negative reinforcement, sufficient wateravailable on the S+ trials alone etc). We speculate that if e.g. the mixture / same source stimulus as well as the individual odours had prior, inherent value / saliency (like e.g. in the moth where specific blends of odours elicit stereotypical behaviours (Martin *et al.,* 2013, PMID 24002682)), learning might be very rapid – and we are currently developingalternative behavioural paradigms to investigate this in more detail.

Most importantly, however, in the context of our manuscript, after mice had learned to perform the "same source – different source" task, they naturally and without any training classified correlated stimuli reliably and as accurately as animals trained on thespecific pulse trains.

Appendix

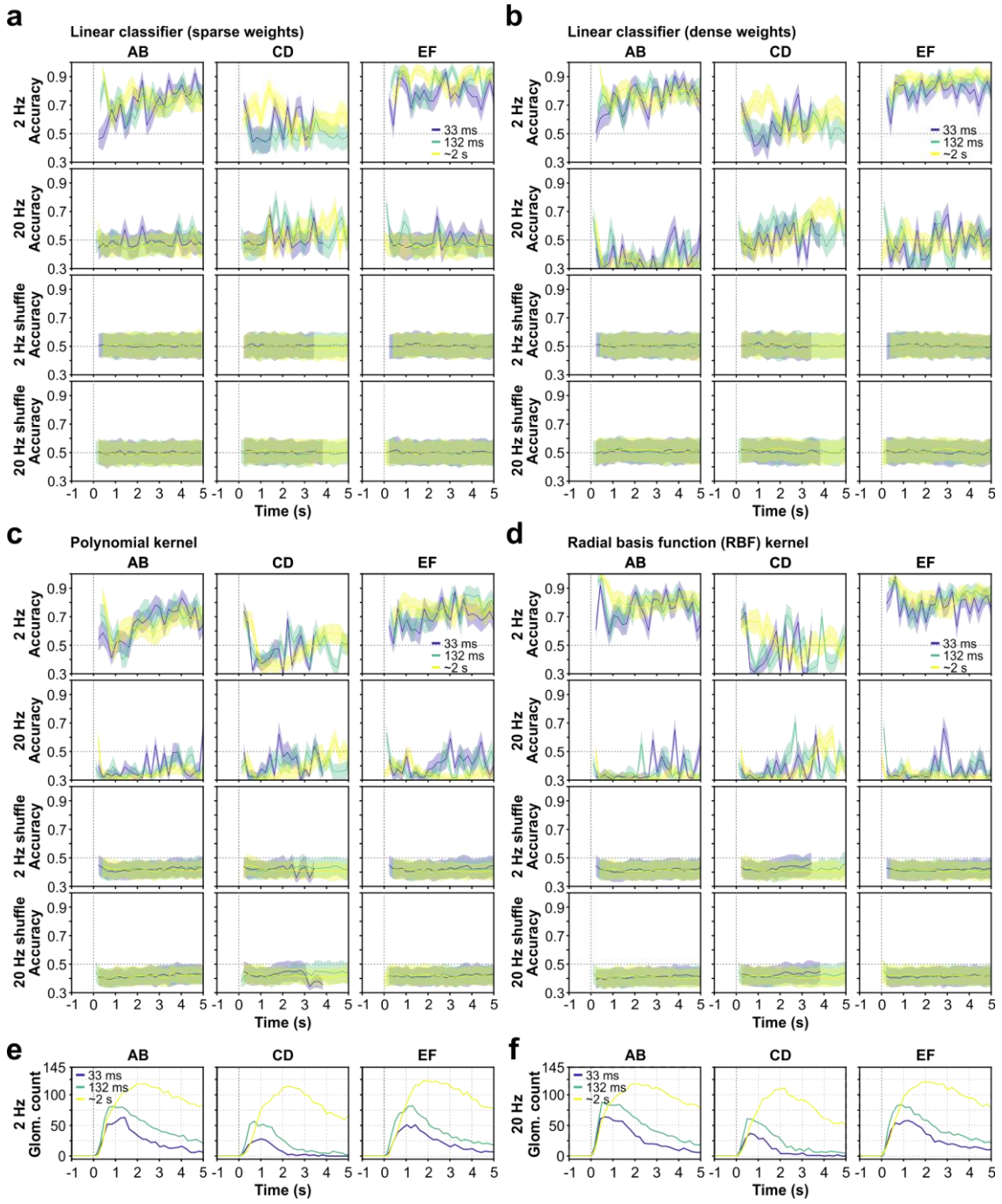

Figure R21 - Appendix: Classifiers. a, Decoding performance of a sparsity-promoting linear classifier (the Lasso evaluated as a classifier; see Methods) using the activity of all responsive ROIs (see panels **e**,**f** for counts, Methods for definition of responsivity), averaged over the specified time bins (colours), and spaced every 200 ms. X-coordinates indicate latest extent of each window relative to odour onset. Shown are mean (lines) ±1 SD (bands) of classifier accuracy computed over 100 random seeds. The accuracy for each seed is defined as the mean accuracy over 10 random cross- validation trials. Patchiness is because not all time points had responsive ROIs for all time windows. Vertical line indicates odour onset. Horizontal line indicates chance level. **b,** As in **a** but for a support vector classifier with l2 penalty, which tends to learn dense decoding weights. Performance is similar to **a** but shows some overfitting at 20 Hz, in particular to odour AB. **c**, as in **a,b** but for a non-linear decoder, a support vector classifier with cubic kernel. Performanceis worse than in previous panels and shows considerable overfitting. **d,** as in the previous panels but for another non-linear decoder, a support vector classifier with a radial-basis-function kernel. Performance at 2 Hz is comparable to the sparse decoders, but overfits elsewhere. **e**, the number of responsive ROIs (see Methods for definition, for each time point (x-coordinate)

nature portfolio

and window (colours) and odour pair (panels), at 2 Hz. **f,** as in **e** but for 20 Hz. Full code and data to produce this figure can be found at https://github.com/stootoon/crick-osn-decoding-release.

Reviewer Reports on the First Revision:

Referee #2 (Remarks to the Author):

The authors have answered all my questions. This is an outstanding contribution to the understanding of how sensory systems deal with complex input.

Referee #3 (Remarks to the Author):

The authors have responded extensively and adequately to the previous questions/concerned I raised. I have no further concerns.

Author Rebuttals to First Revision (please note that the authors have quoted the reviewers in blue text and responded in black text):

Referees' comments:

Referee #2 (Remarks to the Author):

The authors have answered all my questions. This is an outstanding contribution to the understanding of how sensory systems deal with complex input.

Thank you very much for your support and constructive comments throughout the reviewing process!

Referee #3 (Remarks to the Author):

The authors have responded extensively and adequately to the previous questions/concerned I raised. I have no further concerns.

Thank you very much for your support and constructive comments throughout the reviewing process!