Contribution of behavioural variability to representational drift

Sadra Sadeh[1*] and Claudia Clopath[1*]

[1] Bioengineering Department, Imperial College London, London SW7 2AZ, United Kingdom

*Correspondence: C.C. (c.clopath@imperial.ac.uk), or S.S. (s.sadeh@imperial.ac.uk)

Abstract

Neuronal responses to similar stimuli change dynamically over time, raising the question of how internal representations can provide a stable substrate for neural coding. While the drift of these representations is mostly characterized in relation to external stimuli or tasks, behavioural or internal state of the animal is also known to modulate the neural activity. We therefore asked how the variability of such modulatory mechanisms can contribute to representational drift. By analysing publicly available datasets from the Allen Brain Observatory, we found that behavioural variability significantly contributes to changes in stimulus-induced neuronal responses across various cortical areas in the mouse. This effect could not be explained by a gain model in which change in the behavioural state scaled the signal or the noise. A better explanation was provided by a model in which behaviour contributed independently to neuronal tuning. Our results are consistent with a view in which behaviour modulates the low-dimensional, slowly-changing setpoints of neurons, upon which faster operations like sensory processing are performed. Importantly, our analysis suggests that reliable but variable behavioural signals might be misinterpreted as representational drift, if neuronal representations are only characterized in the stimulus space and marginalised over behavioural parameters.

Introduction

Neuronal responses to stimuli, contexts or tasks change over time, creating a drift of representations from their original patterns[1–6]. This representational drift can reflect the presence of intrinsic noise or plasticity in the circuitry and, depending on its origin, can be detrimental to or beneficial for the neural code[7,8]. Understanding the mechanisms contributing to the emergence of representational drift can therefore shed light on its relevance for neural computation[2,8].

Representational drift can arise from bottom-up mechanisms, like changes in the feedforward input to neurons or from a dynamic reorganization of recurrent interactions in the network. Another important source of variability that can contribute to representational drift is changes in the behavioural state of the animal. Spontaneous behaviour has in fact been shown to heavily modulate responses in awake behaving animals[9–11]. Drift of behavioural state – e.g. changes in attention, arousal or running – can therefore change the way neural activity is modulated by top-down mechanisms[9,12] over different timescales.

The exact manner in which such top-down mechanisms modulate the neural activity[13–17] would in turn determine how the behavioural drift affects the representational drift. One possibility is that stimulus-evoked responses are just scaled by arousal or running, as suggested by gain models[18]. Under this scenario, the behavioural state of the animal can modulate the similarity of sensory representations across multiple repeats of the same stimulus (representational similarity), by increasing or decreasing the signal-to-noise ratio. Another possibility is that the behaviour contributes independently to neuronal activity, and hence representational similarity is better described in a parameter space where internal and external parameters conjointly define the neural code. Under the latter scenario, variability in behavioural "signal" could be perceived as noise from the viewpoint of sensory representations, and could therefore be mistaken as representational drift.

To delineate the contribution of behavioural variability to representational drift and to shed light on the involved mechanisms, we analysed publicly available datasets from the Allen Brain Observatory[19,20]. First, we found that behavioural variability strongly modulates similarity of neuronal representations in response to multiple repeats of the same stimulus. In fact, our results suggest that a significant fraction of what has been described as representational drift in a sensory cortex can be attributed to behavioural variability. Second, we found evidence for independent contribution of behaviour to neuronal responses. Our analysis suggests that the contribution of external and internal parameters to representational similarity would be better understood when representations are described in a multidimensional parameter space which is not marginalised over behavioural parameters.

Results

Representational similarity depends on the behavioural state of the animal

We analysed publicly available, large-scale, standardized *in vivo* physiology datasets recently published by the Allen Brain Observatory[19]. The electrophysiology data obtained via Neuropixels probes[21] provides the possibility of studying the spiking activity of a large number of units to visual stimuli (see Methods). We studied similarity of neural activity in response to multiple repeats of the same natural movie (Fig. 1). This was quantified by a measure of representational similarity, which was characterized as the similarity of responses, at the population level, to multiple repeats of the stimulus (Methods and Extended Data Fig. 1).

The analysis was performed in two datasets with different structure of stimulus presentations (Fig. 1a,e; see Supplementary Table 1). In each dataset, the natural movie (30 second long) is presented multiple times in each block of presentation (10 and 30 repeats for dataset1 and dataset2, respectively). We analysed the data for two blocks of presentation separated by more than an hour (Fig. 1a,e). For each presentation of the natural movie, we calculated a population vector of responses from the average activity of all neurons in the primary visual cortex (V1), in bin widths of 1 second starting from the onset of movie presentation (Methods). Representational similarity between two repeats of the natural movie was quantified by the correlation coefficient of the population vectors (Extended Data Fig. 1c and Methods).

Previous analysis has shown that such representational similarity is higher within a block of presentation, and decreases significantly between different blocks, both in Neuropixels and calcium imaging datasets[5]. Our results confirmed this observation, but we also found that representational similarity is strongly modulated by the behavioural state of the animal. This was most visible in sessions where the behavioural state (as assayed by pupil diameter and the average running speed) changed clearly between the two repeats of the movie (Fig. 1a,e). We observed that, firstly, change in the behavioural state strongly reduced the representational similarity between the two blocks (Fig. 1b,f), reminiscent of the representational drift which has been reported over the scale of hours to days[4–6]. Secondly, increased pupil diameter and running during the second block of presentation in fact increased the similarity of responses to the same movie within that block (Fig. 1b,f, left). Overall, there was a significant drop of representational similarity between the movie repeats in which the animal experienced the most changes in the average pupil size (Fig. 1b,f, right). These results indicate that the behavioural state of the animal can bidirectionally modulate the representational similarity across repeats of the same stimulus.

We found similar dependence of representational similarity on the pupil change for other sessions (Fig. 1c,g) and across all animals (Fig. 1d,h). The effect was more prominent when focusing on movie repeats with significant changes in the average running (Fig. 1d,h, left, black lines). Similar trend was also observed when considering units from all recorded regions, instead of only focusing on V1 units (Extended Data Fig. 2a). We also observed the same trend when repeating the analysis within blocks (Fig. 1d-h, right, grey lines, and Extended Data Fig. 2), although the drop of representational similarity across blocks was more prominent due to more drastic behavioural changes between the blocks, which is expected from the slow timescale of changes in the behavioural state.

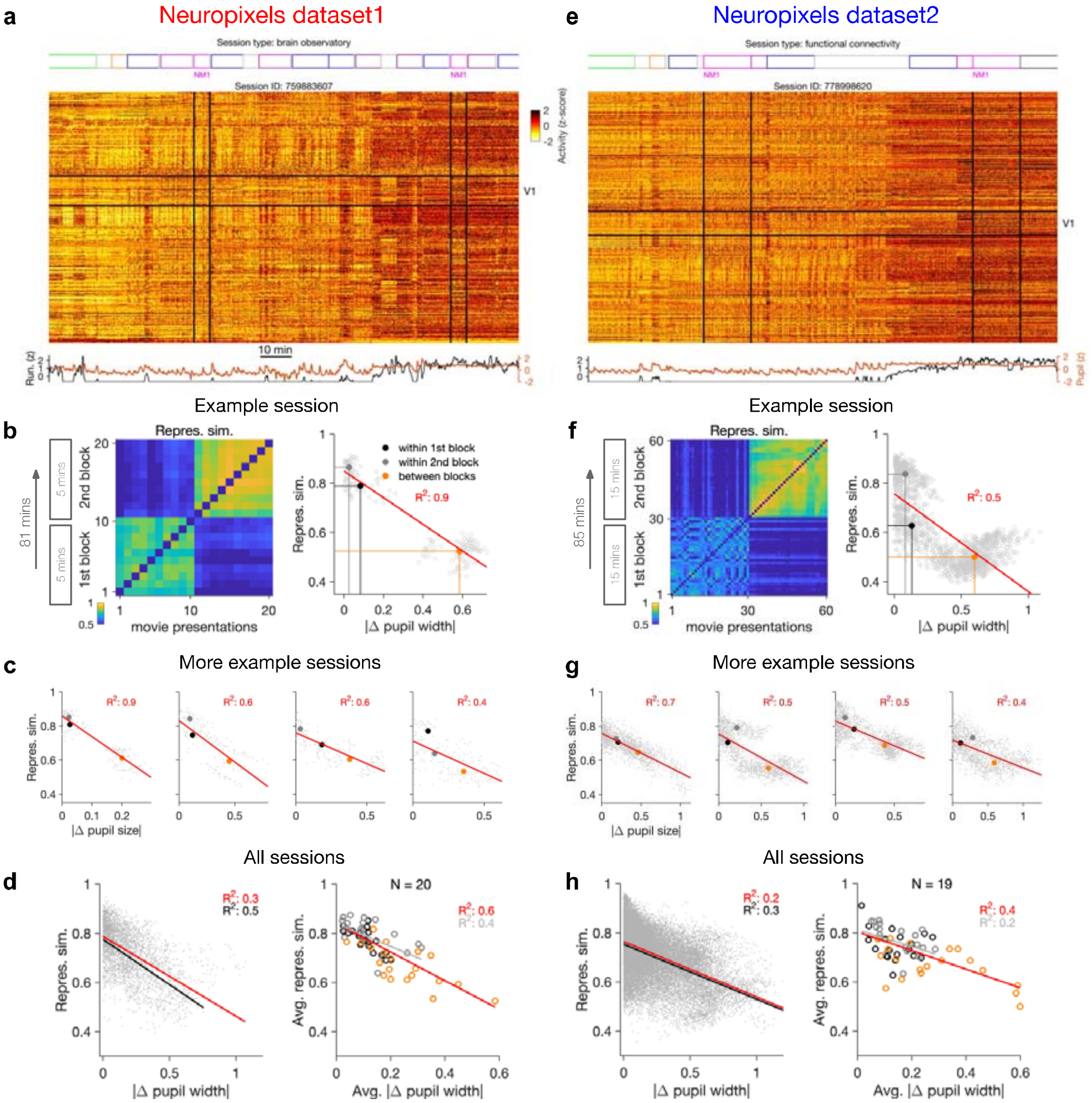


Fig. 1: Representational similarity depends on the behavioural state of the animal.
a, Responses of units measured in an example session to different stimuli denoted on the top. Spiking activity of units is averaged in bins of 1 second and z-scored across the entire session for each unit.

Units in primary visual cortex (V1) and the two blocks of presentation of natural movie 1 (NM1) are highlighted by the black lines. Pupil size and running speed of the animal (z-scored) shown on the bottom. **b**, Representational similarity between different presentations of natural movie 1. It is calculated as the correlation coefficient of vectors of population response of V1 units to movie repeats (see Methods). Left: The matrix of representational similarity for all pairs of movie repeats within and across the two blocks of presentation. Right: Representational similarity as a function of the pupil change, which is quantified as the normalized absolute difference of the average pupil size during presentations (see Methods). The best fitted regression line (using least squares method) and the R squared value (R^2) are shown. Filled circles show the average values within and between blocks. **c**, Same as (**b**, right) for four other example recording sessions. Session numbers and the number of V1 units (#) in each case, respectively: 762602078 (#75), 750332458 (#63), 760345702 (#72), 751348571 (#49). Only sessions with #>40 are included in the analysis. **d**, Same as (**b**, right) for all recording sessions. Left: Data similar to (**c**, grey dots) are concatenated across all mice and the best fitted regression line to the whole data is plotted. Black line shows the fit when movie repeats with significant change in the average running speed of the animal is considered (80th percentile). Right: The average values within and between blocks (filled circles in (**c**)) are plotted for all mice and the fitted regression line to these average values is plotted. Grey lines and R^2 values indicate the fit to within-block data only. N: number of mice. **e-h**, Same as (**a-d**) for a different dataset. Session numbers and the number of V1 units (#) in (**g**), respectively: 766640955(#52), 787025148(#68), 771990200(#54), 829720705(#52).

In the above analysis, we considered the actual spiking activity of the units to build the population vectors. Calculating the representational similarity from these vectors can potentially bias the estimate by increasing the impact of highly active neurons. For instance, if the units which are firing higher remain consistently active, they may lead to some similarity of population vectors even independent of stimulus-evoked responses. To control for variance in the average activity of units, we repeated our analysis for population vectors composed of z-scored responses (as shown in Fig. 1a,e; see Methods). Overall, representational similarity diminished when calculated from the z-scored activity (Extended Data Fig. 2b). However, we observed the same trend in terms of dependence on the behavioural state, whereby larger changes in pupil size were correlated with larger changes in representational similarity (Extended Data Fig. 2b).

Our previous analyses were performed on wild-type mice as well as mice from three different transgenic lines (Pvalb-IRES-Cre × Ai32, n=8; Sst-IRES-Cre × Ai32, n=12; and Vip-IRES-Cre × Ai32, n=8; see Supplementary Table 1)[19]. Inclusion of multiple cell types may distort our estimate of representational similarity, especially as different cell classes can be differentially modulated by behaviour[14,16]. To control for this, we repeated our analysis for recording sessions in wild type mice only and observed similar results (Extended Data Fig. 3a). Our results also held when analysing female and male animals separately (female mice comprised a smaller fraction of the datasets; ~20%) (Extended Data Fig. 3b). Taken together, these results suggest that, in awake behaving animals, variability of the behavioural state is an important factor contributing to the modulation of representational similarity.

Evidence for independent modulation of responses by stimulus and behaviour

What is the mechanism by which behavioural state modulates representational similarity? Changes in pupil area are correlated with the level of arousal[22], which can modulate the neuronal gain[17]. We therefore studied a possible gain model in which changes in pupil size modulate the neuronal responses to sensory inputs (Fig. 2a; Methods). Alternatively, rather than scaling the stimulus-induced signal, behaviour can contribute independently to neuronal activity[16,23]. We therefore compared the gain model to a model in which the neural tuning was obtained by an independent mixing of stimulus and behavioural signals (Fig. 2b; Methods).

In each model, we calculated representational similarity from the neuronal responses to presentations of the same stimulus, and plotted that against the relative behavioural parameter (B) across the repeats (B_i/B_j, for the i-th and j-th repeats) (Fig. 2c,d). Both models showed, on average, a similar dependence of representational similarity on relative behaviour (Fig. 2c,d; the gain model only showed the same pattern if the signal was scaled by the behaviour; we observed different patterns, if behaviour scaled the noise, or both the signal and the noise; Extended Data Fig. 4a,b).

To compare different models with the experimental results, we took the relative pupil size as a proxy for relative behaviour and plotted the representational similarity of all units against it (Fig. 2e). This revealed a similar average dependence as the signal-gain model and the independent-mixing model (Fig. 2c-e). We observed a similar dependence for both datasets, and for most individual recording sessions within each dataset (Extended Data Fig. 4c-f). Similar results were observed when representational similarity was calculated from V1 units or all recorded units (Extended Data Fig. 4c-f).

We then asked how, at the level of individual units, the modulations of responses by stimulus and behaviour are related to each other (Fig. 2f,g). To this end, instead of calculating representational similarity at the population level, we quantified the similarity of individual units' responses to multiple repeats of the stimulus (stimulus reliability; see Methods and Extended Data Fig. 1d). In the signal-gain model, stimulus reliability was highly correlated with behavioural modulation of units (Fig. 2f). This is a consequence of the scaling of the signal by the behaviour, which implies that neurons with higher signal component also show higher modulation with the behavioural parameter (see Methods). The signal-gain model therefore predicts that neurons which are strongly modulated by the stimulus also show strong modulation by the behaviour (Fig. 2f). In contrast, the independent-mixing model predicted an independent relationship between stimulus and behavioural modulation of individual units (Fig. 2g).

We tested these predictions in experimental data, by calculating behavioural modulation and stimulus reliability of individual units in all mice across both datasets. Behavioural modulation was calculated as the correlation of each unit's activity with pupil size, and stimulus reliability was obtained as the average correlation of each unit's activity vectors across multiple repeats of the natural movie (Methods and Extended Data Fig. 1d). As opposed to the signal-gain model, we did not observe a correlation between stimulus and behavioural modulation (Fig. 2h). In fact, a regression analysis suggested that the two modulations are independent of each other in both datasets (Fig. 2h), consistent with the independent-mixing model.

Overall, there was a wide distribution of stimulus reliability (Extended Data Fig. 5a) and behavioural modulation (Extended Data Fig. 5c) across recorded units, with patterns highly consistent across the two datasets. Most V1 units showed variable responses to repeats of the natural movie, as indicated by the peak of the distribution at low stimulus reliability (Extended Data Fig. 5a). However, the distribution had a long tail composed of units with high stimulus reliability, which showed highly reliable responses across repeats of the movie (Extended Data Fig. 5a,b). There was a wide spectrum of behavioural modulation too, with most units showing positive correlations with pupil size (Extended Data Fig. 5c,d), and a smaller population of negatively modulated units (Extended Data Fig. 5c).

The units that showed significant modulation with the stimulus were not necessarily modulated strongly with the behaviour parameter, and vice versa; in fact, it was possible to find example units from all four combinations of weak/strong x stimulus/behavioural modulations (Extended Data Fig. 5e,f). A clear example of the segregation of stimulus and behavioural modulation was observed in CA1, where the units showed, on average, very weak stimulus reliability across movie repeats, consistently across different mice and datasets (Extended Data Fig. 6a). However, they were largely modulated by behaviour, to an extent comparable to V1 units (Extended Data Fig. 6a-c). Taken together, these results suggest that, rather than simply scaling the stimulus-evoked responses, behaviour modulates the activity in a more independent and heterogeneous manner.

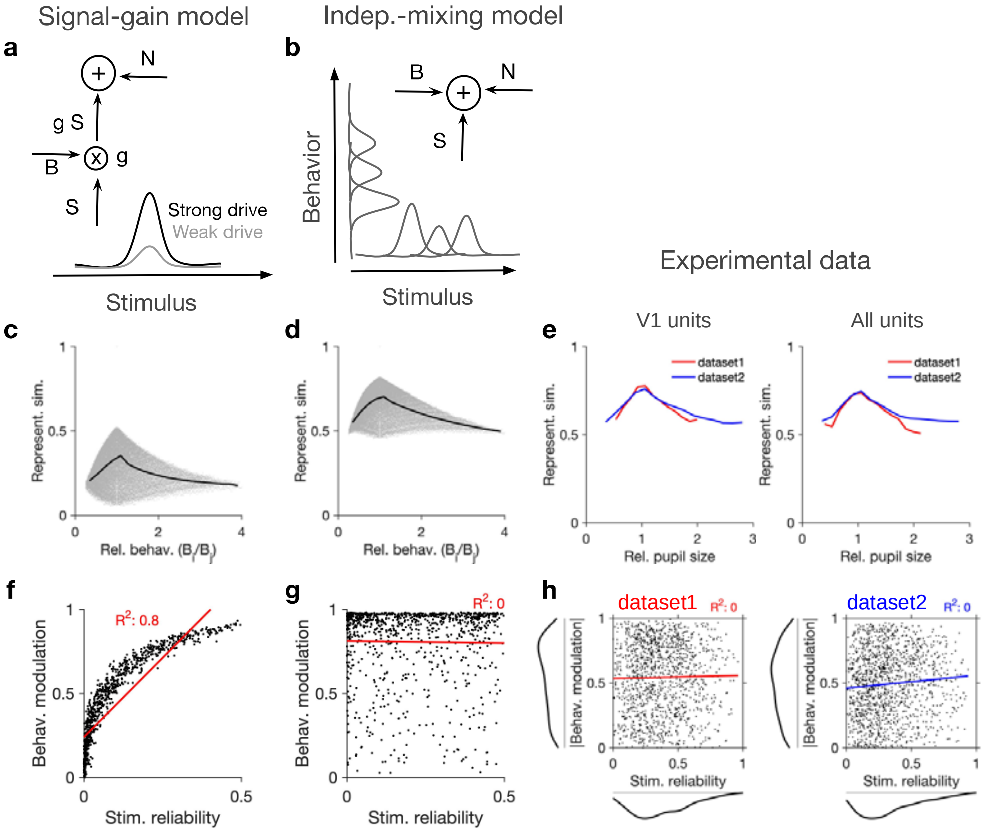


Fig. 2: Independent modulation of activity by stimulus and behaviour.

a, Schematic of a signal-gain model in which behaviour controls the gain with which the stimulus-driven signal is scaled. Individual units are driven differently with the stimulus, leading to different tuning curves which determines their stimulus signal, S. Behavioural parameter, B, sets the gain, g, with which the stimulus signal is scaled, before being combined with the noise term, N, to give rise to the final response. S is the same across repetitions of the stimulus, while N is changing on every repeat (see Extended Data Fig. 4a and Methods). **b**, An alternative model (independent-mixing) in which the response of a unit is determined by the summation of its independent tuning to stimulus and behaviour (Methods). **c**, Representational similarity of population responses to different repeats of the stimulus as a function of the relative behavioural parameter (B_i/B_j) in the signal-gain model. Black line shows the average (in 20 bins). **d**, Same as c, for the independent-mixing model. **e**, Same as (c,d) for the experimental data from Neuropixels dataset1 (red) or dataset2 (blue). For each pair of movie repeats, the average representational similarity of population responses (left: V1 units; right: all units) are plotted against the relative pupil size (P_i/P_j). (**f,g**) Relation between behavioural modulation and stimulus reliability of units in different models. The stimulus signal-gain model predicts a strong dependence between behavioural modulation and stimulus reliability of units (f), whereas the independent-mixing model predicts no correlation between the two (g). Best fitted regression lines and R^2 values are shown for each case. **h**, Data from V1 units in the two datasets show no relationship between the stimulus reliability of units and their absolute behavioural modulation, as quantified by the best fitted regression lines and R^2 values. Stimulus reliability is computed as the average correlation coefficient of each unit's activity vector across repetitions of the natural movie, and behavioural modulation is calculated as the correlation coefficient of each unit's activity with the pupil size (see Methods). Marginal distributions are shown on each axis.

Behavioural variability modulates the low-dimensional components of population activity independent of stimulus reliability

If the behavioural state of the animal modulates the neuronal responses independently of the stimulus, it should be possible to see its signature in the low-dimensional space of neural activity. To test this, we analysed the principal components (PCs) of population responses in individual sessions (Fig. 4). For the two example sessions we analysed previously (shown in Fig. 1a,f), the first two PCs explained a significant fraction of variance (Fig. 4a,g). Low-dimensional population activity showed a distinct transition between two behavioural states, which were corresponding to low versus high arousal, as quantified by different pupil sizes (Fig. 4b,h). The first PC, which explained most of the variance was, in fact, strongly correlated with both pupil size and running speed (Fig. 4c,i). These results suggest that behavioural modulation contributes significantly to the low-dimensional variability of neural activity.

To link the low-dimensional modulation of activity by behaviour to single neurons, we next analysed the projection of units' activity over the PCs. PC projections in the neural space indicated a spectrum of weakly- to highly-active units (Extended Data Fig. 7a). In fact, neural projections over the first two PCs were correlated with the average activity of neurons (Extended Data Fig. 7b). In contrast to the average activity, the PC projections did not reveal any relationship with the stimulus reliability of units (Fig. 4d,j), suggesting that the low-dimensional neural activity is modulated independently of stimulus-evoked responses. These results were remarkably consistent across different datasets and across difference mice (Fig. 4e,f,k,l). The first two PCs explained similar levels of variance across more than 20 mice in each dataset (Fig. 4e,k). In both datasets, the regression analysis revealed no relationship between the two PC projections and the stimulus reliability of units (Fig. 4f,l; see Extended Data Fig. 7c for individual sessions). We therefore conclude that behaviour significantly modulates the low-dimensional components of neural activity, but this modulation does not specifically enhance the activity of neurons which are more reliably representing the stimulus.

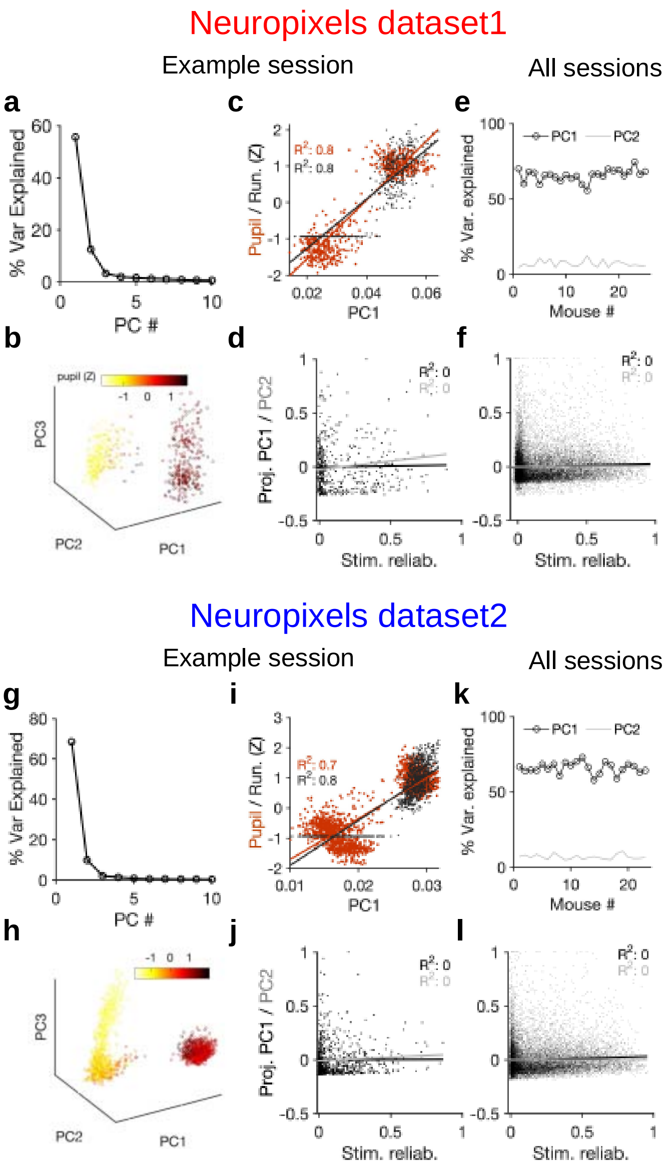


Fig. 3. Behavioural variability modulates the low-dimensional components of population activity independent of stimulus reliability.

a, Relative contribution of the first 10 principal components (PCs) of population responses to the variability of activity (quantified by the fraction of explained variance by each PC) for an example session (same as shown in Fig. 1a). **b**, Population activity in the low-dimensional space of the first three PCs. Pseudo colour code shows the pupil size at different times, indicating that the sudden transition in the low-dimensional space of activity is correlated with changes in the behavioural state. **c**, Strong correlation between PC1 and the behavioural state of animal (assayed by either pupil size or running speed). **d**, Projection of units' activity over PC1/PC2 versus stimulus reliability of the unit reveals no correlation between the two (as quantified by best fitted regression lines in each case). The best fitted regression lines and R^2 values in each case are shown. **e**, Fraction of variance explained by the first and second PCs for all sessions. **f**, Same as (**d**) for all units from all sessions in dataset1. **g-l**, Same as (**a-f**) for dataset2.

Behaviour modulates the setpoint of responses

To gain further insight into how the behaviour modulates the low-dimensional pattern of population activity, we analysed the relation between behavioural parameters and the average activity of units. In the two example sessions analysed previously (shown in Fig. 1), there was a transition in the average pupil size and running speed in the second block, which was correlated with an overall increase in the average population activity (Fig. 5a,e and Extended Data Fig. 8a). In general, change in the pupil size explained a significant fraction of changes in population activity of V1 units in all sessions (Extended Data Fig. 8b).

We also looked at the average activity of individual units across all movie frames and repetitions (their setpoints). Units had a wide range of setpoints, which were relatively stable within each block (small variance across repetitions relative to the mean) (Extended Data Fig. 8c). However, the setpoints changed upon transition to the next block, with most units increasing their setpoints, without an apparent link to their previous setpoint levels (Extended Data Fig. 8d). The population vectors composed of setpoints in each repeat can be further used to quantify setpoint similarity (Fig. 5b,f). Within-block correlations were high, indicating the stability of setpoints when behavioural changes were minimum – although occasional, minor changes of pupil size still modulated these correlations (Fig. 5b,f). Most changes in setpoint similarity, however, appeared between the blocks, when the animal experienced the largest change in its behavioural state.

Quantifying the dependence of setpoint similarity on changes in pupil size revealed a strong relationship, both for V1 units and for all recorded units (Fig. 5c,g). The relationship was rather stable when calculated from responses to single frames of movie presentation, instead of the average activity across the entire movie (Extended Data Fig. 8e). We obtained similar results when the dependence was calculated from the average block activity across all mice, from both datasets (Fig. 5d,h). These results, therefore, suggest that the behavioural signal can modulate the setpoint of neural activity independent of stimulus, and, in doing so, induce a similarity (/dissimilarity) of population responses when behaviour is stable (/changing).

Note that an unintuitive connotation of this finding is that quantifying response similarity from population vectors (see e.g.[5]) may reveal "representational drift" upon behavioural changes, even independent of stimulus-evoked modulation of activity. This is because the constancy of setpoint activity of units would lead to some degree of similarity between population vectors, even if the stimulus-evoked component is different (Extended Data Fig. 6d). The behaviourally-induced component of similarity changes more slowly, leading to a drop in representational similarity on a longer timescales (e.g. between blocks of stimulus presentation, rather than within them). In line with this reasoning, we observed a similar drop of "representational similarity" in CA1 (Extended Data Fig. 6e), although individual units in this region had, on average, no reliable visual representations (Extended Data Fig. 6a). Modulation of setpoint activity – and hence setpoint similarity – by the behaviour can, therefore, contribute to representational similarity, independent of stimulus-evoked responses.

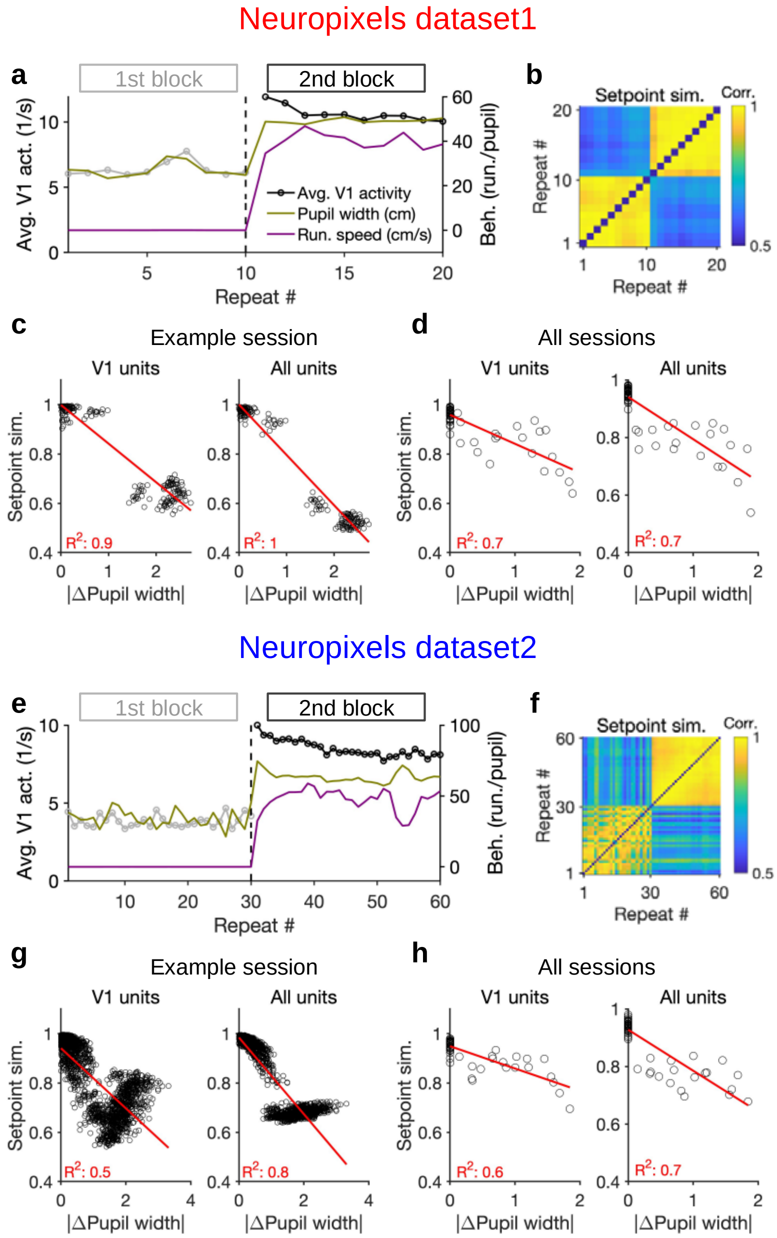


Fig. 4: Behaviour modulates the setpoint of responses. **a**, Average population activity and behavioural parameters (pupil size and running speed) during the first and second blocks of presentation of natural movie 1 (same examples sessions as Fig. 1). Grey, first block; black, second block; each point corresponding to the average in one repeat of the movie. **b**, Setpoint similarity is calculated as the correlation coefficient of population vectors composed of average activity of units during each repeat of movie presentation. Change in the behavioural state (as quantified by the pupil size) between the two blocks is correlated with a drastic decrease in the average between-block setpoint similarity. Note that transient changes of pupil within each block also modulate the setpoint similarity. **c**, Setpoint similarity (as in **b**) as a function of change in pupil size between the movie repeats, when the population vectors of setpoints are composed of V1 units (left) or all recorded units (right). **d**, Dependence of setpoint similarity on pupil change for all sessions, calculated from within-block and across-block averages in each session. **e-h**, Same as (**a-d**) for dataset2.

Behaviour reliably modulates responses during active states

What distinguishes an independent behavioural signal from a modulatory component or from noise is that it brings about reliable responses for different states of behaviour. That is, there should exist a reliable and independent tuning of units with behavioural parameters (like pupil size or running speed). We therefore investigated more precisely how the neural activity is modulated by behaviour (Fig. 6). We used the correlation of units' activity with running as a metric for behavioural tuning. To obtain a measure of significance of correlations, we calculated bootstrapped correlations (see Methods). More than half of the units showed significant modulation by running, and the fraction and distribution of significant correlations were similar between the two blocks and across the two datasets (Extended Data Fig. 9).

Another way to assay the reliability of behavioural tuning is to test whether the correlation of units with behaviour remains stable between the two blocks of presentation (Fig. 6a,d). Random correlations with running should be uncorrelated across the two repeats. In contrast, regression analysis revealed a good correlation between the two blocks (Fig.6a,d, left). The distributions of correlations with behaviour were also similar between the two blocks (Extended Data Fig. 9). Notably, focusing on sessions with similar levels of running between the two blocks (Fig. 6b,e), and on units with significant behavioural modulation, improved the similarity of tuning between the two repeats (Fig. 6a,d, right). Specifically, most units which were positively (/negatively) modulated during the first block remained positively (/negatively) modulated in the second block (Fig. 6a,d, right). These results therefore suggest that a significant fraction of the population shows reliable modulation by running – similar result is expected for pupil, as we observed a high correlation between modulation of units with running and pupil in both datasets (Fig. 6b,e, lower).

Our results held when repeating the analysis for all units instead of V1 units only (Fig. 6c,f and Extended Data Fig. 9). We also observed similar results when quantifying the reliability of tuning between two blocks of presentation of another stimuli (drifting grating; Extended Data Fig. 10). Notably, the tuning of units remained stable from one stimulus type to another: modulation of units during presentation of drifting gratings had a good correlation with their modulation during natural movie presentations for both blocks (Extended Data Fig. 10d,h). The tuning with running was even reliable between the first (30-90 mins) and second (90-150 mins) parts of the entire session, with each part containing different stimuli (Extended Data Fig. 11). We did a region-specific analysis of this reliability and found that reliable tuning exists in various regions (Extended Data Fig. 11). Overall, these analyses suggest that behaviour reliably and independently modulates neuronal responses.

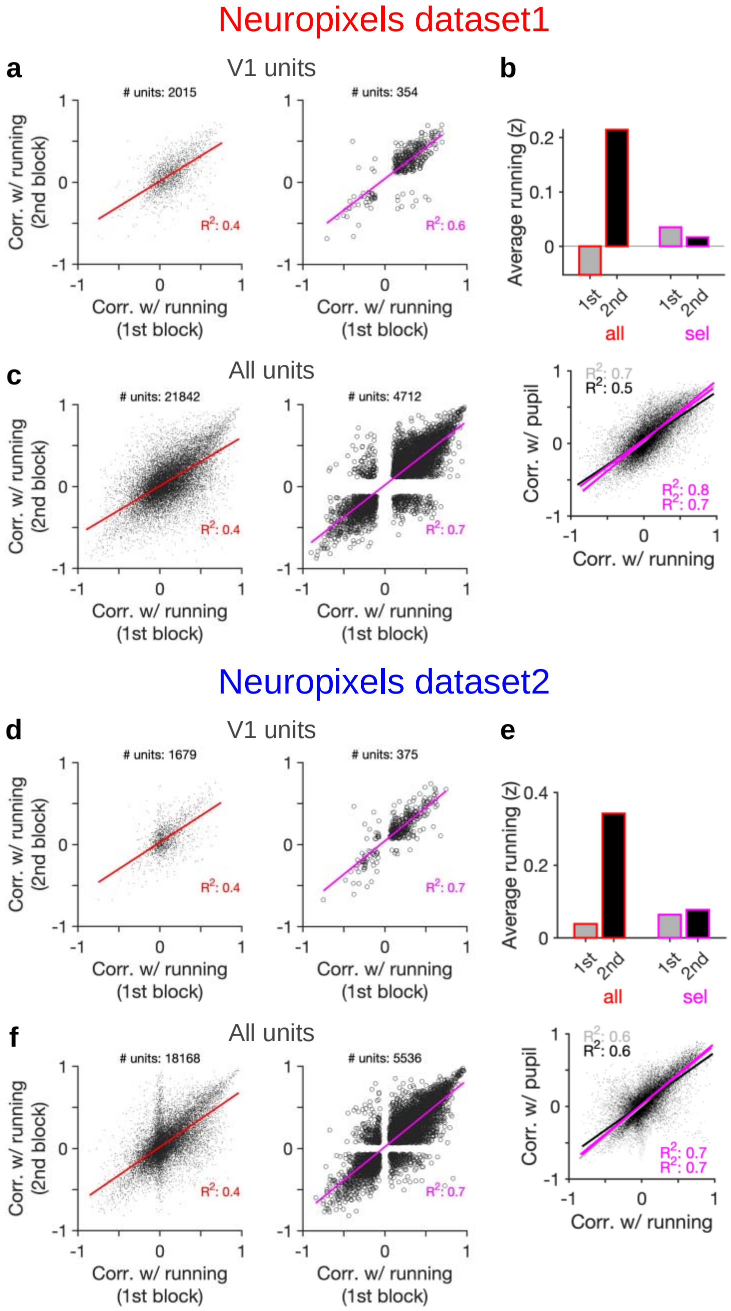


Fig. 5: Behaviour reliably modulates responses during active states.

a, Correlation of activity with running during second block against the first block for all V1 units (left), and for selected sessions and units (right). In the latter case, only sessions with similar average running between the two blocks, and units with significant correlation with running, are selected (see Extended Data Fig. 9 and Methods for details). **b**, Upper: Average z-scored value of running in the first (1st) and the second (2nd) block across all units/sessions (all; red) and for selected ones (sel; magenta). Lower: Correlation of all V1 units with pupil size against their correlation with running in first (grey) and second (black) blocks. Magenta: regression fits for selected units/sessions only. **c**, Same as (**a**) for recorded units from all regions. **d-f**, Same as (**a-c**) for dataset2.

Stimulus-dependence of behavioural variability and setpoint similarity

External stimulus directly modulates the responses by activating selective receptive fields of neurons, which can be measured under anaesthesia[24,25]. In awake behaving animals, however, it is possible that different stimulus types indirectly modulate the responses by inducing different patterns of behavioural variability. We indeed found that this was the case when comparing natural movies with an unnatural stimulus (drifting gratings) (Fig. 6). Natural movies induced more variability of pupil size and running in the animals across the two blocks of stimulus presentations: both measures significantly increased during the second block for natural movies, whereas changes were not significant for drifting gratings (Fig. 6a,d). The result was consistent across the two datasets with different length and sequence of stimulus presentations (cf. Extended Data Fig. 11a,b).

To see if and how this difference affects response similarity of units, we calculated average setpoint similarity (cf. Fig. 5) between the two blocks of presentations from the shuffled activity of units in response to different stimuli (see Methods). Average setpoint similarity was high for both stimuli, but it was significantly larger for drifting gratings for most sessions (Fig. 6b,e). Plotting setpoint similarity as a function of behavioural changes for the entire distribution revealed its composition across the two stimulus types. Responses to drifting gratings showed, on average, a higher setpoint similarity for similar behavioural states (small behavioural changes) (Fig. 6c,f), arguing for more stability of average responses even independent of behavioural variability. Larger behavioural changes were more prevalent for the natural movie presentations, and units' responses showed a large drop of setpoint similarity at these deviations (Fig. 6c,f), leading to a significant drop of average setpoint similarity compared to drifting gratings. Taken together, these results suggest that stability of population responses to different stimulus types might be determined by the combined effect of stimulus-evoked reliability of responses and its indirect modulation by behavioural variability.

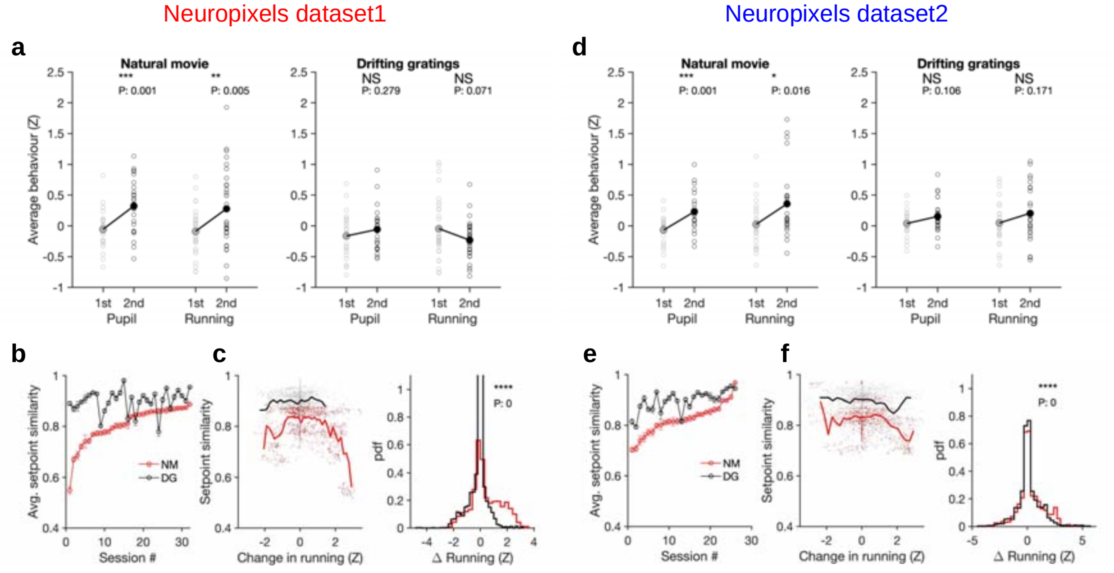


Fig. 6: Stimulus-dependence of behavioural variability and setpoint similarity.

a, Average pupil size and running speed during the 1st (grey) and 2nd (black) blocks of presentation of natural movies (left) and drifting gratings (right) for different sessions (empty circles). Filled circles: the mean across sessions. Pupil size and running speed are z-scored across each session, respectively. P-values on top show the result of two-sample t-tests between the two blocks. NS: P > 0.05. *: P ≤ 0.05; **: P ≤ 0.01; ***: P ≤ 0.001; ****: P ≤ 0.0001. **b**, Average setpoint similarity between the two blocks of presentation of natural movie 1 (NM) and drifting gratings (DG) for different sessions. Sessions are sorted according to their average setpoint similarity for NM. Population vectors are built out of the average responses of all units to 30 randomly chosen frames (1 second long). The correlation coefficient between a pair of population vectors from different blocks (within the same stimulus type) is taken as setpoint similarity. The procedure is repeated for 100 pairs in each session and the average value is plotted. Error bars show the std across the repeats. **c**, Left: Setpoint similarity as a function of the difference in average running, ΔZ=Z2-Z1, where Z1 and Z2 are the average running during randomly chosen frames in the 1st and 2nd block, respectively. The lines show the average of the points in 40 bins from the minimum ΔZ to the maximum. Right: Distribution of changes in running for different stimuli. **d-f**, Same as (**a-c**) for dataset2.

Decoding generalizability improves by focusing on reliable units

How does behavioural variability affect the decoding of stimulus-related information, and how can decoding strategies be optimized to circumvent the drift of representations? Our analyses so far suggested that behaviour modulates the responses in addition to and independently of stimulus-evoked modulations (independent model in Fig. 3). This independent behavioural modulation would be perceived as noise, if a downstream decoder is decoding stimulus-related signals, and can compromise the generalizability of decoding. For instance, the activity of a subpopulation of units (A) might be assigned to stimulus A by the decoder, in the absence of significant behavioural modulation. If the decoder is now tested in a new condition where behaviour modulates the responses independently of the presented stimulus, the activity of subpopulation A can be interpreted as presence of stimulus A, independent of the presented stimulus. This is in contrast to the gain model (signal-gain model in Fig. 2b) in which behavioural state scales the stimulus-evoked signal, and can therefore not compromise the generalizability of decoding (subpopulation A only responds to stimulus A, but with different gains). In the signal-gain model, focusing on units which are strongly modulated by behaviour should in fact enhance the decoding generalizability under behavioural variability, whereas in the independent model the focus should be on units with more stimulus reliability.

We tested these two alternatives directly by training a linear decoder to discriminate between different frames of the natural movie (Fig. 7a and Methods). The decoder was trained on the activity of units in the first block to detect a target frame; it was then tested on the second block of presentation to discriminate between the target frame and other frames, in order to evaluate the generalizability of decoding (i.e out-of-distribution transfer) (Fig. 7a). When the decoder was trained on the activity of all units in the first block, discriminability (d') was very low in the second block (Fig. 7b,c,e,f). However, focusing on the reliable units (units with high stimulus reliability) shifted the distribution of d' to larger values and increased the average discriminability (Fig. 7c,f). Focusing on units with strong behavioural modulation, on the other hand, did not yield higher discriminability in the second block (Extended Data Fig. 12). These results suggest that behavioural modulation is detrimental to generalizability of stimulus decoding, and that this problem can be circumvented by focusing on units with more stimulus information.

This effect was consistent across mice in both datasets (Fig. 7d,g). In dataset2, we observed higher average d' in the second block, when the decoder was trained and tested on all units. This could be due to more presentations of the natural movie in dataset2 (30 repetitions in each block compared to 10 in dataset1). Larger training samples can help the decoder in learning the signal from the noise, suggesting that the effect of behavioural "noise" on corrupting the stimulus signal is more significant for small sample sizes. On the other hand, longer presentations can lead to sampling from responses under more behavioural variability, which can in turn inform the decoder on how to ignore the stimulus-irrelevant modulation by behaviour. Altogether, these results corroborate our previous analysis that the contribution of behavioural variability to neural activity is orthogonal to stimulus modulations, and suggest that such behavioural noise limits the decoding capacity especially with limited data.

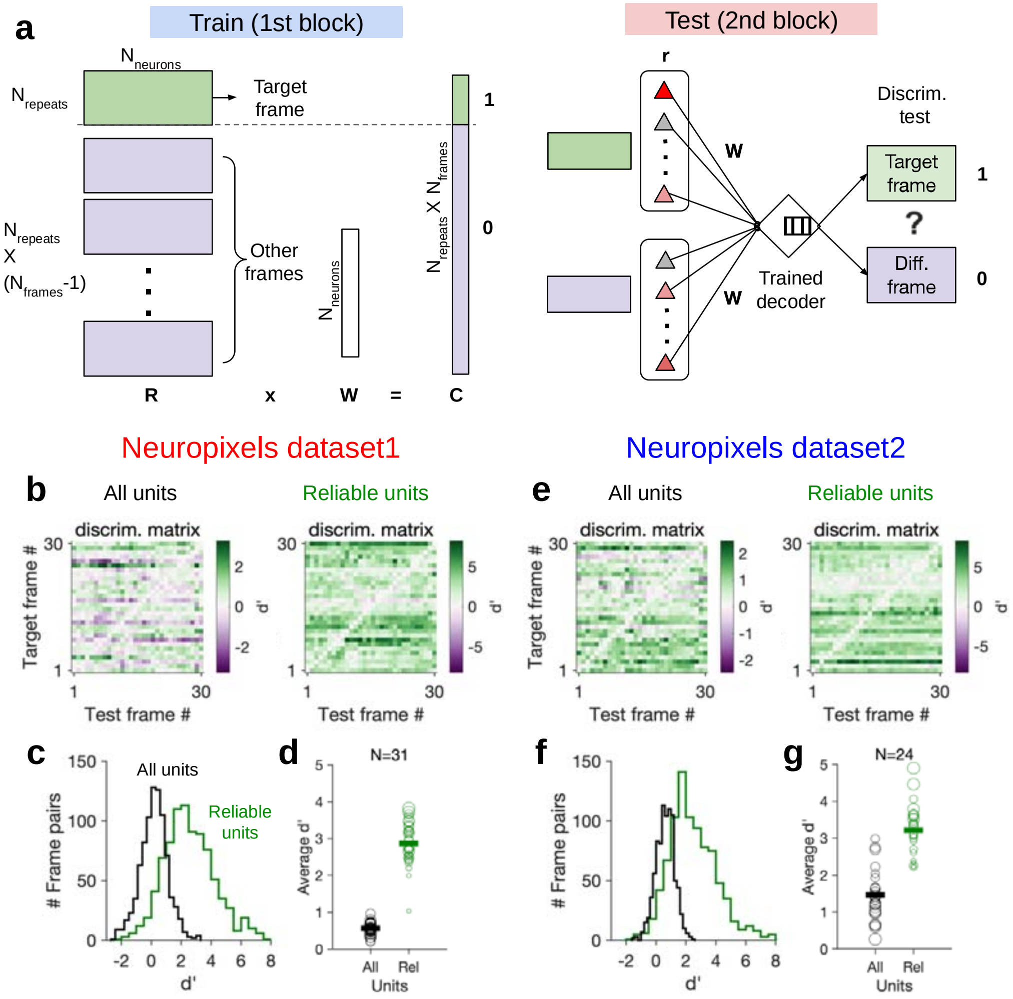


Fig. 7: Decoding generalizability of natural images improves by focusing on reliable units.
a, Schematic of a decoder which is trained on the population activity during the first block of presentation of the natural movie (upper) and tested on the second block of presentation to discriminate different frames of natural images from each other (out-of-distribution transfer, see Methods for details). **b**, Matrix of discriminability index (Methods), d', between all combination of movie frames as target and test, when all units (left) or only units with high stimulus reliability (right) are included in training and testing of the decoder. **c**, Distribution of d' from the example discriminability matrices shown in (**b**) for decoders based on all units (black) and reliable units (green). Reliable units are chosen as units with stimulus reliability (Methods) of more than 0.5. **d**, Average d' for all mice, when all units or only reliable units are included. Size of each circle is proportionate to the number of units available in each session (sessions with >10 reliable units are included). Filled markers: average across mice. **e-g**, Same as (**b-d**) for dataset2. Data in (**b,c**) and (**e,f**) are from the same example sessions shown in Fig. 1a and Fig. 1e.

Discussion

The results of our analysis here suggest that variability of the behavioural state of animal contributes significantly to changes in representational similarity. We found that population responses to different repeats of the same natural movie was the most similar when behavioural parameters like pupil size and running speed changed the least. This was a result of an independent modulation of neural activity by behaviour, which was mixed with stimulus-evoked responses to represent a multidimensional code. Our results are consistent with a view in which behaviour modulates the low-dimensional, slowly-changing setpoints of neurons, upon which faster operations like sensory processing are performed.

Small modulation of ongoing neural dynamics by sensory stimuli was reported before in awake, freely viewing animals[26], in line with other reports on the significance of internal signals even in sensory cortices[27–29]. Our results here are consistent with these reports, and our analysis provides a mechanism by which variability of the internal state can contribute to ongoing signal correlations. It suggests that two distinct sources of response similarity exist in neuronal networks, with one set by baseline responses modulated on a slower timescale via internal parameters (setpoint similarity), and the other arising from finer and faster modulations invoked by sensory stimuli. Importantly, changes in representational similarity (i.e. representational drift) can arise from changes in both sources, and hence attributing it purely to the drift of the sensory component might be inaccurate.

Internal, behavioural states of the animal can contribute independently to neural processing, or can act as a modulator for external stimuli, for instance by increasing the input gain and enhancing the saliency of the sensory signal. Notably, our results could not be explained by a model in which behaviour acted as a gain controller for sensory inputs. Such a model would predict a direct relationship between the stimulus modulation and behavioural modulation of neurons. One would therefore expect that the most reliable neurons in representing sensory information to be modulated the most by arousal or running states. However, we found that the reliability of stimulus-evoked responses to different repeats of the same natural movie was independent of behavioural modulation, in line with previous reports[16].

A gain-model account of behavioural modulation would only change the signal-to-noise ratio of sensory representations by behaviour. Therefore, if the level of arousal or attention of the animal drifts away over time, the signal component of the representations becomes weaker compared to the noise, leading to some drop in representational similarity. In contrast, independent modulation of neuronal responses by behaviour affects representational similarity in more complex ways. First, similarity of population vectors across repeats of the same stimuli can be due, at least partly, to the behavioural signal rather than stimulus-evoked responses. Second, changes in behavioural signal might be perceived as sensory-irrelevant noise, if the parameter space of representations (composed of internal and external parameters) is only analysed over the external stimulus dimension[10,30]. Behavioural variability can, therefore, be misinterpreted as representational drift in the latter scenario.

A recent analysis of similar datasets from the Allen Brain Observatory reported similar levels of representational drift within a day and over several days[5]. While this was mainly attributed to the drift of sensory representations, our analysis here shows that the contribution of behavioural variability might have a strong contribution in these (and possibly other) datasets. First, the study showed that tuning curve correlations between different repeats of the natural movies were much lower than population vector and ensemble rate correlations[5]; it would be interesting to see if, and to which extent, similarity of population vectors due to behavioural signal that we found here (cf. Fig. 5) contributes to this. Second, the stimulus-independent component of representational drift due to behavioural variability is a global phenomenon that can affect all regions, even independent of their involvement in the processing of natural images. Similar to[5], we found "representational drift" in many areas, including regions like CA1, although units in this region had no reliable representation of natural images (Extended Data Fig. 5a). Drawing further conclusions about stimulus-dependences of representational drift in visual cortex – and other sensory cortices – thus needs a critical evaluation by teasing apart the contribution of different components.

Another recent study reported stimulus-dependent representational drift in the visual cortex, whereby responses to natural images experienced large representational drift over weeks compared to responses to drifting gratings[6]. In line with the finding of this study, we found here that responses to drifting gratings were more robust to behavioural variability in general. However, we also observed that different stimulus types can induce variable patterns of behaviour, thus highlighting the combined contribution of behaviour and stimulus to representational drift. Notably, the mentioned study[6] found a dependence of representational drift on the pupil size (see Supplementary Fig. 8c in [31]), with a more decrease in pupil size over time correlating with more representational drift for both stimulus types (see Supplementary Fig. 11d in [6]). Such consistent changes of behaviour may contribute to representational drift over longer timescales (days to weeks), by recruiting similar mechanisms as we described here for shorter intervals (e.g. changes in setpoint similarity). Mapping behavioural changes over longer times and per individual animal can shed light on the specific contribution of behaviour to representational drift. It would for instance be interesting to see if the large variability of representational drift across different animals (see Supplementary Fig. 5 in the same study[6]) might be linked to their behavioural variability.

Behavioural variability might be more pertinent to other modalities for which active sensing is less constrained during experiments. While eye movements are minimized in head-fixed mice, in other modalities (like olfaction) it might be more difficult to control for the behavioural variability arising from active sensing (e.g. sniffing) over time and across animals. A recent study demonstrated significant representational drift over weeks in the primary olfactory cortex of mouse[4]. The surprising finding that sensory representations are not stable in a sensory cortex was hypothesized to be linked to the different structure of piriform cortex compared to other sensory cortices with more structured connectivity. It would be interesting to see if, and to which extent, other factors like changes in the gating of olfactory bulb by variable top-down modulations[32,33], or changes in the sniffing patterns of animals, may contribute to this. Similar to the general decline over time of the pupil size reported in the visual cortex[6], animals may change their sniffing patterns during experiments, which can in turn lead to a general or specific suppression or amplification of odours, depending on the level of interest and engagement of individual animals in different sessions.

Beyond sensory processing, variability of internal state can also contribute to other cognitive processes in various cortices[34]. A recent study in monkey found that changes in the perceptual behaviour was modulated by a slow drift in its internal state, as measured by pupil size[35]. This was correlated with a slow drift of activity in V4 and PFC, along with changes in the impulsivity of the animal (as reflected in the hit rates), which overrode the sensory evidence. These results, in another species, are in agreement with our findings here on the contribution of behavioural drift to changes in neural representations. Interestingly, the sensory bias model in the study could not capture the effect of the slow drift on decoding accuracy; instead, an alternative impulsivity model, which introduced the effect of slow drift as an independent behavioural parameter, matched with the data (Fig. 6 in [35]).

Another study in monkey M1 found that learning a new BCI task was modulated along the dimension of neural engagement of the population activity, which in turn was correlated with pupil size[36]. Neural engagement increased abruptly at the beginning, and decreased gradually over the course of learning, where output-null and output-potent components of neural engagement differentially attuned for different targets. Notably, exploiting behavioural perturbations in this study enabled an interactive interrogation of the neural code during learning. Behavioural perturbations, combined with large-scale recording and perturbation of neural activity[37–39], which are more feasible in mice, can pave the way for a more precise (and potentially causal) interrogation of the neural mechanisms underlying representational drift. It would specifically be interesting to see how the bidirectional modulation of activity by behaviour we observed here emerges and which circuit mechanisms[13,14,17,18] contribute to it.

In summary, our analysis reveals new insights on representational drift from the viewpoint of behaviour. Conceptually, it argues for primacy of internal parameters[40], and suggests that representational similarity could be better understood and characterized in a multidimensional parameter space where the contribution of external and internal parameters are equally considered. Computationally, it argues for an independent mixing of stimulus-evoked and behavioural signals, rather than a simple gain modulation of sensory inputs by behaviour. Technically, it asks for further controls and analysis of behavioural variability in the characterisation of representational drift. Future studies will hopefully probe the multidimensional code underlying representations in the brain by combining large-scale recordings of neural activity with simultaneous measurement and quantification of behaviour.

Methods

Curation and preprocessing of the data

Data curation. Publicly available data provided by the Allen Brain Observatory[19,20] was accessed via AllenSDK (https://allensdk.readthedocs.io). We analysed recording sessions in which neuronal responses to visual stimuli were measured via electrophysiology techniques by Neuropixels probes (https://portal.brain-map.org/explore/circuits/visual-coding-neuropixels). The data composed of 58 sessions/mice in two separate datasets: brain observatory dataset (Dataset1; n=32) and functional connectivity (Dataset2; n=26) (Supplementary Table 1). Similar stimuli (including natural moves and drifting gratings) were shown to the animals, with different length and sequence of presentations in each dataset (https://allensdk.readthedocs.io/en/latest/_static/neuropixels_stimulus_sets.png; see Extended Data Fig. 11a,b for illustration of different stimulus sets). We used the spiking activity of units which was already extracted by Kilosort2[10], and we included units in our analysis which passed the default quality criteria. Invalid intervals were treated as Not a Number (NaN) values. For further details on the preparation of animals, visual stimulation, data acquisition and default pre-processing of data, see the Technical White Paper from the Allen Brain Observatory on Neuropixels Visual Coding.

Pre-processing of data. For our analysis here, we rendered the spiking activity of units in bins of 1 second. When analysis was focused on specific stimulus types (e.g. presentation of natural movie 1 as in Fig. 1b,f), the activity was rendered from the onset of presentation of each block of the stimulus. When the analysis was across all stimuli and involved the activity during the whole session (e.g. data shown in Fig. 1a,e), the activity was rendered from the beginning of the session or an arbitrary time (e.g. time frames specified in Extended Data Fig. 11). Behavioural information was obtained in similar time frames. Locomotion was quantified for all animals as the average running speed. Level of arousal was quantified by pupil size, as measured by pupil width (whenever pupillometry was available; Supplementary Table 1).

To normalize the parameters (e.g. to normalize for different size of pupil across animals), we calculated their z-score values. For parameter x (units' activity, pupil size or running speed), it was obtained as $z = (x - \mu_x)/\sigma_x$, where μ_x and σ_x are the mean and standard deviation of x during the entire session or a specified time window.

Data analysis

Representational similarity. Representational similarity of population activity was quantified by calculating the correlation of responses to different repeats of the same stimulus (Extended Data Fig. 1c). Let v be a vector of responses of N recorded units to M 1-second long chunks of a natural movie (the natural movie is broken down to M chunks, or frames, each lasting for 1 second, corresponding to the bin width the neural activity is rendered in). v is a $1 \times \mathrm{N}M$ population vector composed of the concatenated activity of units (either the actual activity, i.e. average spiking activity, or the z-scored activity of each unit). Denote v_i and v_j as vectors of responses to two repeats of the same natural movie. Representational similarity is quantified as the Pearson correlation coefficient of these two population vectors:

$$\rho_{ij} = \frac{\mathrm{cov}(v_i, v_j)}{\sigma_{v_i}\sigma_{v_j}}$$

Stimulus reliability. We also quantified the reliability of how single units respond individually to repetitions of the stimuli (Extended Data Fig. 1d). To quantify that, we calculated a stimulus reliability metric, which is obtained as the average correlation coefficient of each unit's activity vector (r) across repetitions of the stimulus (e.g. the natural movie). Let ${r^k}_i$ and ${r^k}_j$ be the vectors of response of the k-th unit to the i-th and j-th repetitions of the natural movie. Similarity of the unit's response between these two repeats can be quantified by the Pearson correlation coefficient of the responses as before:

$$\rho_{ij}^k = \frac{\mathrm{cov}\left(r_i^k, r_j^k\right)}{\sigma_{r_i^k}\sigma_{r_j^k}}$$

Stimulus reliability of the unit k is obtained as the average correlation across all pairs of (non-identical) repetitions of the stimulus:

$$\rho^k = \frac{1}{N_r(N_r - 1)} \sum_{i=1}^{N_r} \sum_{j \neq i} \frac{\mathrm{cov}\left(r_i^k, r_i^k\right)}{\sigma_{r_i^k}\sigma_{r_i^k}}$$

where N_r is the number of repetition of the stimulus. Note that to each single unit we can ascribe a stimulus reliability index, since this is calculated from the individual vectors of single units' responses (r^{k}); on the other hand, representational similarity is calculated from the population vector of responses (v) and indicates a single population metric ascribed to the activity of a group of neurons (e.g. V1 units or all recorded units).

Behavioural tuning. To obtain a measure of how single units are modulated by behaviour, we calculated the correlation of units' responses with behavioural parameter, β:

$$\rho_i(\beta) = \frac{\mathrm{cov}(r_i, \beta)}{\sigma_{r_i}\sigma_\beta}$$

Here, r_i is the vector of response of the i-th unit, and β is the vector of respective behavioural parameter (either pupil size or running speed) rendered during the same time window and with the same bin width as unit's activity.

To obtain a measure of reliability of this modulation by behaviour, we calculated bootstrap correlations. The activity of each unit was shuffled for 100 times and the correlation with behaviour was calculated. The mean (μ_{sh}) and std (σ_{sh}) of the distribution of shuffled correlations were then used to obtain the z-scored, bootstrapped correlation:

$$Z = \frac{\rho(\beta) - \mu_{sh}}{\sigma_{sh}}$$

where $\rho(\beta)$ is the unshuffled correlation of the unit with behaviour.

Modelling

Gain models. To gain mechanistic insight on the contribution of behavioural changes to modulation of representational similarity, we explored two different models. First, we developed a gain model, in which the integration of the signal and the noise by neurons was differently modulated by behaviour (Extended Data Fig. 4a). For a population of N_p neurons, let u be the $1 \times N_p$ vector of responses of neurons upon presentation of a stimulus. This is assumed to be composed of signal (S) and noise (N) components. Change in the behavioural parameters (for instance, pupil size) is supposed to change a gain parameter, g, which in turn differently modulate the signal (S) and noise (N). The vector of population activity, u, is obtained, as a linear combination of weighted components by the behavioural/gain parameter. If the signal and the noise are both scaled by the behavioural parameter, it is given as $u = \mathrm{gS} + \mathrm{gN}$. If either the noise or the signal are scaled, it is given as $u = \mathrm{S} + \mathrm{gN}$ or $u = \mathrm{gS} + \mathrm{N}$, respectively (Extended Data Fig. 4b).

S and N are both vectors of size $1 \times N_p$, where each element is drawn from a random uniform distribution between $[0,1]$. The population activity is simulated for N_r repeats of the stimulus. The stimulus signal, S, remains the same for all the repeats (frozen noise drawn from the same range as before, $[0,1]$), while the noise component, N, is instantiated randomly on each repeat (from the same range, $[0,1]$, as the signal). The behavioural parameter (e.g. pupil size) is assumed to change on every repeat too, which changes the gain parameter, g, as a result. g was therefore assumed to be a random number uniformly drawn from the range $[0.5, 2]$ for each repeat. We chose $N_p = 1000$ and $N_r = 100$.

Representational similarity for different models was calculated, similar to the procedure in analysing the experimental data, as:

$$\rho_{ij} = \frac{\mathrm{cov}(u_i, u_j)}{\sigma_{u_i}\sigma_{u_j}}$$

where u_i and u_j are population responses to the i-th and j-th repeat of the stimulus, obtained from different gain models. This value is plotted against the relative gain (obtained as the ratio of the gains in the two repeats, g_i/g_j or g_j/g_i) in Extended Data Fig. 4b.

Extended gain model. To match better with the experimental data on a single unit level, we extended the previous signal gain model to have stimulus tuning for individual units (Fig. 2a). Whereas before the stimulus was assumed to be a single, fixed value between $[0,1]$ for each neuron, now the stimulus itself is extended (corresponding to different frames of the natural movie or different orientations of drifting gratings). The stimulus, s, is assumed to be a vector of fixed random values between [0,1] with size $1 \times N_s$. Each neuron, k, has a different stimulus drive/tuning, T_k, with which the stimulus vector is multiplied. T is a vector of size $1 \times N_p$ (number of neurons in the population), randomly drawn from $[0,1]$. Response of the k-th neuron to each repeat of the stimulus is composed of its stimulus signal ($S = T_k s$), which is multiplied by the behavioural gain (g), and an added noise term (N), which is independently drawn for each stimulus and repeat from the range $[0,1]$. $N_s = 10$, $N_p = 1000$, $N_r = 200$.

Independent model. We also developed an alternative model, whereby the effect of behaviour on population responses was modelled as an independent signal (Fig. 2b). Here, instead of scaling signal or noise components of the input, behaviour enters as an independent parameter:

$$u = S_S + N + S_B$$

where S_S and S_B are stimulus-evoked and behavioural signals and N is the noise. S_S and N were instantiated as before, while S_B was determined based on two factors. First, the behavioural parameter, β, which was changing on every repeat, and was simulated, similarly as the behavioural gains before, by a random number between $[0.5, 2]$ for each repeat. Second, the vector of tuning (T_B) of different neurons in the population with the behavioural parameter, which was modelled as a random number between $[0,1]$ for each neuron. The behavioural signal was obtained as: $S_B = \beta T_B$. Representational similarity was computed as before for the population vectors and plotted against the relative behavioural parameters.

Decoding model. To directly compare the stimulus-induced information available in different blocks of stimulus presentation, we developed a decoding model (Fig. 7a). A linear decoder is trained on the neural activity (composed of the average activity of units in response to different repeats of the natural movie) during the first block of presentation to discriminate different frames (1 second long) of the natural movie (Fig. 7a, upper). The weights of the readout (W) for each target frame were optimized to maximize its classification (C=1) against other, non-target frames (C=0). The decoder is then tested on the data in the second block (Fig. 7a, lower). The population activity in response to each frame (the vector of average responses of neurons to a single frame across different repeats) is passed through the decoder to determine whether it is the target (D=1) or not (D=0). Performance of the decoder is quantified by calculating the discriminability (d') as

$$d' = \frac{\mu_s - \mu_n}{\sqrt{\sigma_s^2 + \sigma_n^2}}$$

where μ_s and σ_s are the average and std of D for target frame across repetitions (within the second block), and μ_n and σ_n are similar values for non-target frames. The discrimination matrix (Fig. 7b,e) then shows the discriminability (d') of each movie frame as a target when presented against all other frames.

Theoretical analysis

Gain models. Representational similarity for the responses in the gain models can be calculated as follows.

In the absence of any scaling of the signal or the noise, $u = S + N$, the representational similarity is obtained as the correlation coefficient of responses to a pair of stimulus repeats:

$$\rho_{ij} = \frac{\mathrm{cov}(u_i, u_j)}{\sigma_{u_i}\sigma_{u_j}}$$

where $u_i = \mathrm{S} + N_i$ and $u_j = \mathrm{S} + N_j$. Assuming that S and N have zero means, we can write:

$$\rho_{ij} = \frac{\sigma_{\mathrm{S}}^2}{\sigma_S^2 + \sigma_N^2}$$

where σ_S and σ_N are the std of S and N, respectively. This indicates that representational similarity can be expressed as a function of the relative variability of the signal and the noise. If modulation of the responses due to signal is dominant over the noise, $\sigma_N \gg \sigma_N$, $\rho_{ij} \to 1$.

If both the signal and the noise are scaled by the behavioural parameter, by the gain factor g, as $\mathrm{u = gS + gN}$, we obtain:

$$\rho_{ij} = \frac{g_i g_j\,\sigma_S^2}{\mathrm{g_i g_j}(\sigma_S^2 + \sigma_N^2)} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_N^2}$$

where g_i and g_j are the gains in the i-th and j-th repeat of the stimulus, respectively. Representational similarity, therefore, remains the same under similar scaling of S and N.

If only the noise is scaled by behaviour, we obtain:

$$\rho_{ij} = \frac{\sigma_S^2}{\sqrt{(\sigma_S^2 + g_i^2\sigma_N^2)(\sigma_S^2 + g_j^2\sigma_N^2)}}$$

showing that the larger the gain, the smaller the representational similarity.

Similarly, if only the signal is scaled, representational similarity can be obtained as follows:

$$\rho_{ij} = \frac{\mathrm{g_i g_j}\sigma_S^2}{\sqrt{(\mathrm{g_i^2}\sigma_S^2 + \sigma_N^2)(g_j^2\sigma_S^2 + \sigma_N^2)}}$$

which, if rewritten as:

$$\rho_{ij} = \frac{\sigma_S^2}{\sqrt{(\sigma_S^2 + \sigma_N^2/g_i^2)(\sigma_S^2 + \sigma_N^2/g_j^2)}}$$

shows that larger gains effectively decrease the significance of noise, and hence enhance representational similarity. Specifically, in the limit of very large gains for both repetitions ($g_i \gg 1,\ g_j \gg 1$), we have: $\rho_{ij} \to 1$.

For the specific case where gains are the same between the two repeats ($g_i = g_j = g$), the equation simplifies to:

$$\rho_{ij} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_N^2/g^2}$$

Thus, for similar behavioural states (and hence gains) between the two repeats of the stimulus, representational similarity increases if $g > 1$ and decreases if $g < 1$.

Independent model. For the model in which the stimulus and the behaviour contributes independently to neural responses, representational similarity in response to the same stimulus can be expressed as:

$$\rho_{ij} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_B^2 + \sigma_N^2}$$

where σ_S, σ_B, and σ_N denote the variability of the population response induced by stimulus, behaviour and noise components, respectively. In deriving the above equation, we have assumed that the stimulus and behavioural components of the signal are independent, i.e. $< S_S.S_B >= 0$ (in addition to the noise term being independent of S_S and S_B, respectively). We also assumed that the behavioural signal, $S_B = \beta T$, remained the same between the two repeats (that is, the behavioural parameter was the same: $\beta_i = \beta_j = \beta$. If the behavioural parameter changes between the repeats, the equation can, in turn, be written as:

$$\rho_{\mathrm{ij}} = \frac{\sigma_{\mathrm{S}}^2}{\sqrt{\left(\sigma_S^2 + \beta_i^2\sigma_T^2 + \sigma_N^2\right)\left(\sigma_S^2 + \beta_j^2\sigma_T^2 + \sigma_N^2\right)}}$$

Note that, when representational similarity is only characterized in terms of the stimulus part of the signal, the contribution of behavioural variability is similar to a noise term – decreasing ρ_{ij} for larger values of β. Changes in the behavioural state can, thus, not be distinguished from random variability of the "signal".

Relation between representational similarity and stimulus reliability. As explained above, representational similarity and stimulus reliability are calculated to quantify the similarity of population and single units' responses, respectively, to the repeats of the same stimulus. In fact, representational similarity of a population vector composed of one single unit is the same as the stimulus reliability of that unit. Similarly, if all the units in a population of neurons had the same response profile in response to the stimulus, the stimulus reliability of units would be the same as the representational similarity of the population responses. Although these two measures are related (similar to lifetime sparseness and population sparseness[41]), they are, however, not always directly equivalent to each other.

Consider a single unit, k, which has a constant baseline firing rate of r_b and a component which is modulated by the stimulus, r_m: $r = r_b + r_m$. If the stimulus-modulated component of the response is randomly changing between different repeats of the stimulus, the neuron would have a stimulus reliability of zero: $\rho^k = 0$. A population of units with this behaviour would have an average stimulus reliability of zero. However, the representational similarity of the responses of this population is not necessarily zero. In fact, we may obtain high values of population-level representational similarity, if the baseline component of the responses is significantly larger than their modulation ($r_b \gg r_m$). Under this scenario, representational similarity is calculated from the baseline component of the population responses (v_b), which indeed remains constant across repeats, hence $\rho_{ij} \to 1$.

Reference

1. Ziv, Y. *et al.* Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, (2013).
2. Lütcke, H., Margolis, D. J. & Helmchen, F. Steady or changing? Long-term monitoring of neuronal population activity. *Trends in Neurosciences* **36**, (2013).
3. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. *Cell* **170**, (2017).
4. Schoonover, C. E., Ohashi, S. N., Axel, R. & Fink, A. J. P. Representational drift in primary olfactory cortex. *Nature* **594**, (2021).
5. Deitch, D., Rubin, A. & Ziv, Y. Representational drift in the mouse visual cortex. *Curr. Biol.* (2021). doi:10.1016/J.CUB.2021.07.062
6. Marks, T. D. & Goard, M. J. Stimulus-dependent representational drift in primary visual cortex. *Nat. Commun. 2021 121* **12**, 1–16 (2021).
7. Clopath, C., Bonhoeffer, T., Hübener, M. & Rose, T. Variance and invariance of neuronal long-term representations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, (2017).
8. Rule, M. E., O'Leary, T. & Harvey, C. D. Causes and consequences of representational drift. *Current Opinion in Neurobiology* **58**, (2019).
9. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
10. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science (80-.).* **364**, (2019).
11. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, (2019).
12. Vinck, M., Batista-Brito, R., Knoblich, U. & Cardin, J. A. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron* **86**, 740–754 (2015).
13. Fu, Y. *et al.* A cortical circuit for gain control by behavioral state. *Cell* **156**, 1139–1152 (2014).
14. Pakan, J. M. P. *et al.* Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. *Elife* **5**, (2016).
15. Garcia del Molino, L. C., Yang, G. R., Mejias, J. F. & Wang, X.-J. Paradoxical response reversal of top-down modulation in cortical circuits with three interneuron types. *Elife* **6**, (2017).
16. Dipoppa, M. *et al.* Vision and Locomotion Shape the Interactions between Neuron

Types in Mouse Visual Cortex. *Neuron* **98**, (2018).

17. Cohen-Kashi Malina, K. *et al.* NDNF interneurons in layer 1 gain-modulate whole cortical columns according to an animal's behavioral state. *Neuron* (2021). doi:10.1016/j.neuron.2021.05.001
18. Ferguson, K. A. & Cardin, J. A. Mechanisms underlying gain modulation in the cortex. *Nat. Rev. Neurosci.* **21**, 80–92 (2020).
19. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, (2021).
20. de Vries, S. E. J. *et al.* A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat. Neurosci.* **23**, (2020).
21. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, (2017).
22. Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, (2008).
23. Saleem, A. B., Ayaz, A., Jeffery, K. J., Harris, K. D. & Carandini, M. Integration of visual motion and locomotion in mouse visual cortex. *Nat. Neurosci.* **16**, 1864–1869 (2013).
24. Yoshida, T. & Ohki, K. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nat. Commun.* **11**, 872 (2020).
25. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–36 (2008).
26. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573 (2004).
27. Arieli, a, Sterkin, a, Grinvald, a & Aertsen, a. *Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Science (New York, N.Y.)* **273**, (1996).
28. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954–6 (2003).
29. Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**, 1943–6 (1999).
30. Montijn, J. S., Meijer, G. T., Lansink, C. S. & Pennartz, C. M. A. Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective. *Cell Rep.* **16**, (2016).
31. Marks, T. D. & Goard, M. J. Stimulus-dependent representational drift in primary visual cortex. *bioRxiv* (2020).
32. Boyd, A. M., Sturgill, J. F., Poo, C. & Isaacson, J. S. Cortical Feedback Control of

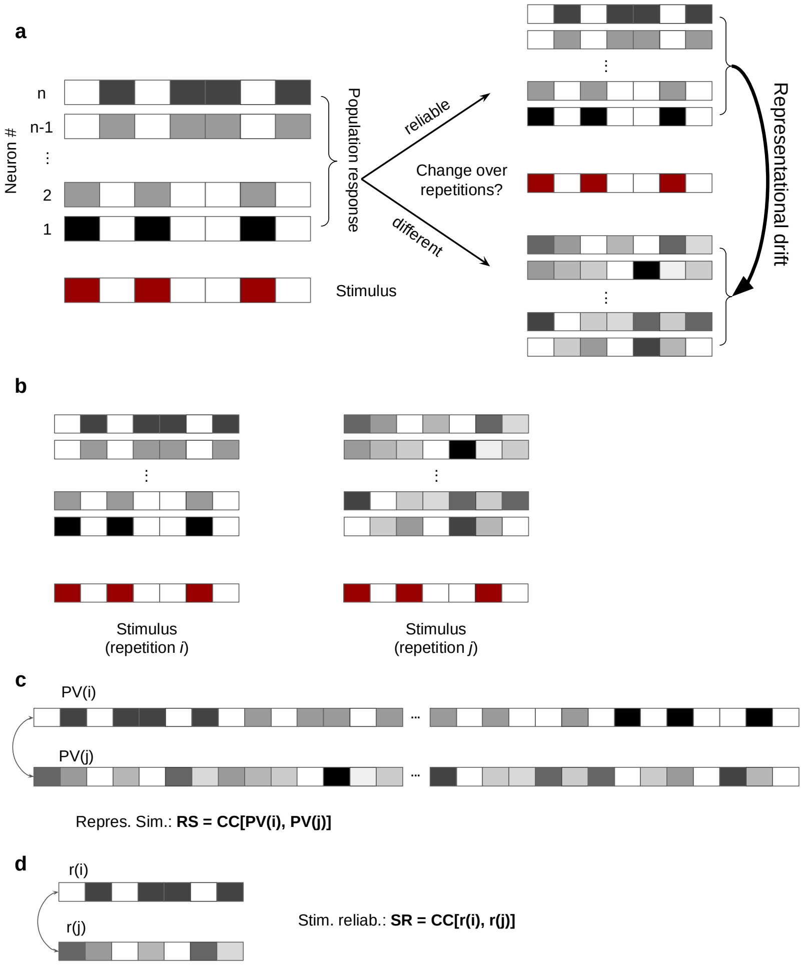
Olfactory Bulb Circuits. *Neuron* **76**, (2012).

33. Markopoulos, F., Rokni, D., Gire, D. H. & Murthy, V. N. Functional Properties of Cortical Feedback Projections to the Olfactory Bulb. *Neuron* **76**, (2012).
34. Joshi, S. & Gold, J. I. Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences* **24**, (2020).
35. Cowley, B. R. *et al.* Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron* **108**, (2020).
36. Hennig, J. A. *et al.* Learning is shaped by abrupt changes in neural engagement. *Nat. Neurosci.* **24**, (2021).
37. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in Neural Systems. *Neuron* **71**, 9–34 (2011).
38. Emiliani, V., Cohen, A. E., Deisseroth, K. & Häusser, M. All-Optical Interrogation of Neural Circuits. *J. Neurosci.* **35**, 13917–13926 (2015).
39. Zhang, Z., Russell, L. E., Packer, A. M., Gauld, O. M. & Häusser, M. Closed-loop all-optical interrogation of neural circuits in vivo. *Nat. Methods* **15**, 1037–1040 (2018).
40. Buzsáki, G. *The Brain from Inside Out*. *The Brain from Inside Out* (2019). doi:10.1093/oso/9780190905385.001.0001
41. Froudarakis, E. *et al.* Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci. 2014 176* **17**, 851–857 (2014).

Acknowledgement

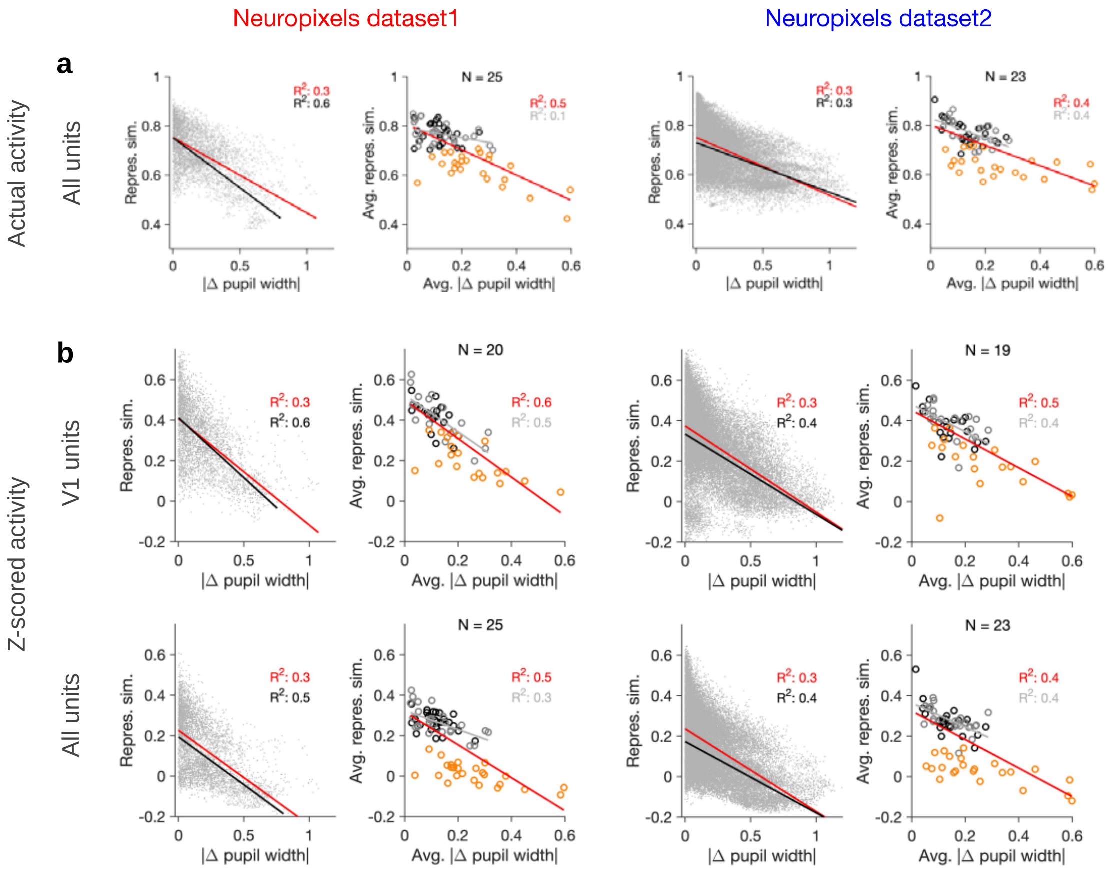
We thank K.A. Wilmes and D.F. Tome for their comments on the manuscript, and all the members of the Clopath Lab for insightful discussions. This work was supported by BBSRC BB/N013956/1, BB/N019008/1, Wellcome Trust 200790/Z/16/Z, Simons Foundation 564408 and EPSRC EP/R035806/1. All data needed to evaluate the conclusions in the paper are presented in the paper and/or the Supplementary Materials.

Supplementary information


a
n
n-1
⋮
2
1
Neuron #
Population response
Stimulus
reliable
different
Change over repetitions?
Representational drift
b
Stimulus (repetition *i*)
Stimulus (repetition *j*)
c
PV(i)
PV(j)
Repres. Sim.: **RS = CC[PV(i), PV(j)]**
d
r(i)
r(j)
Stim. reliab.: **SR = CC[r(i), r(j)]**

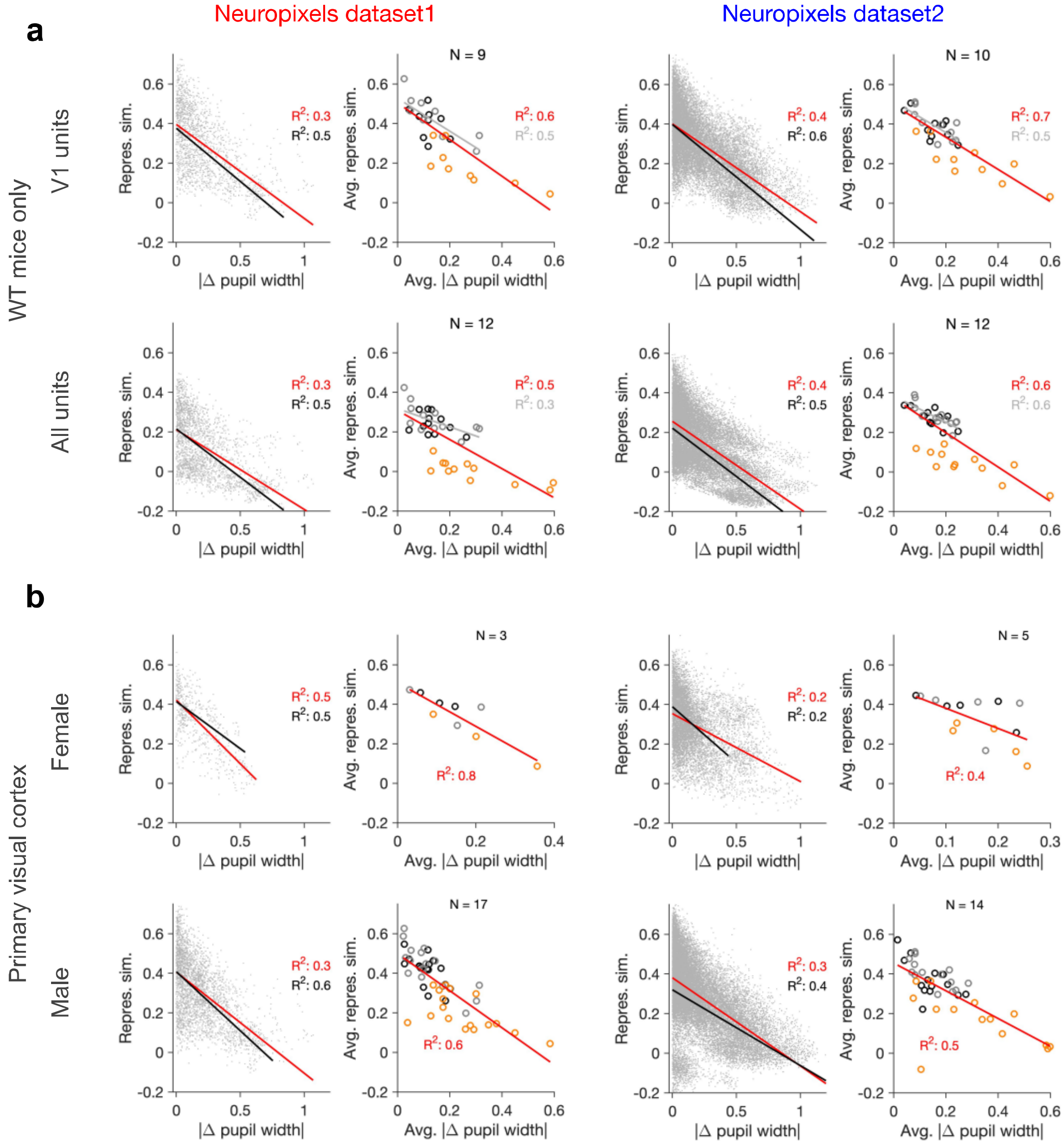
Extended Data Fig. 1: Characterization and quantification of representational similarity.

a, Left: Illustration of response of a population of neurons (#1 to #n; upper) to a stimulus (lower), composed of binary values (ON: red; OFF: white). Right: The population response to another repetition of the same stimulus can remain the same (upper), demonstrating a stable and reliable code, or it can change from the original pattern (lower), leading to a drift of representations. **b**, The degree of change or constancy of representations can be assayed by comparing the population responses to two repeats of the same stimulus. **c**, Representational similarity (RS) is quantified by the corelation coefficient (CC) of the concatenated (across neurons) vector of population responses to two repeats, PV(i) and PV(j). **d**, Stimulus reliability (SR) is calculated for each unit individually, from the CC of the vector of responses of that unit to two stimulus presentations.



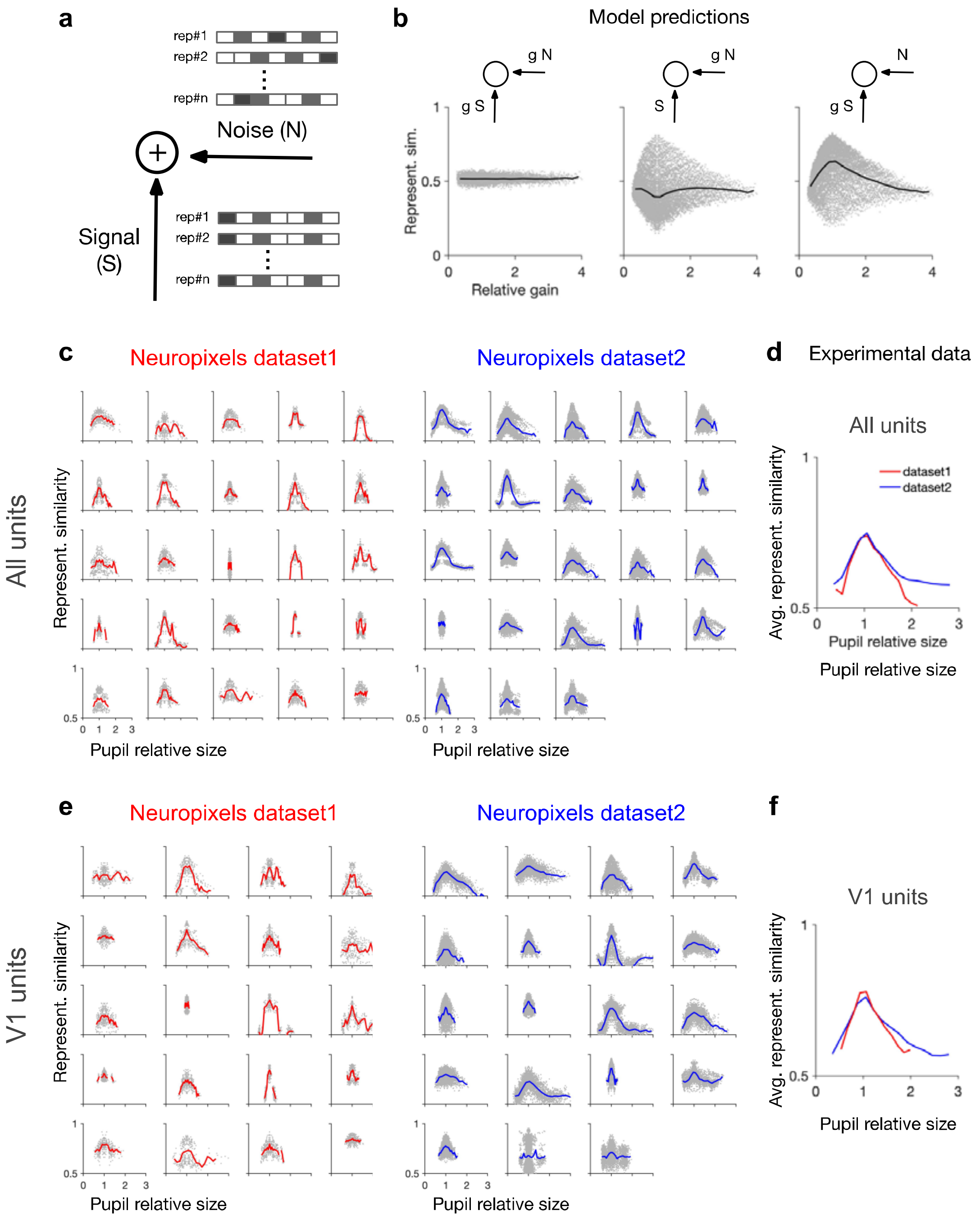
Extended Data Fig. 2: Relation between behavioural changes and representational similarity when calculated from z-scored activity.

a, Same as Fig. 1d,h when all recorded units are included, instead of only V1 units. **b**, Same as Fig. 1d,h when population vectors are composed of z-scored activity of units in V1 (upper) or all regions (lower). Z-scored activity of unit i is calculated as $z_i = (r_i - \mu_i)/\sigma_i$, where μ_i and σ_i are the average and std of the activity of unit (r_i) during the two blocks of presentation of natural movie 1. Left: Neuropixels dataset1; Right: Neuropixels dataset2.



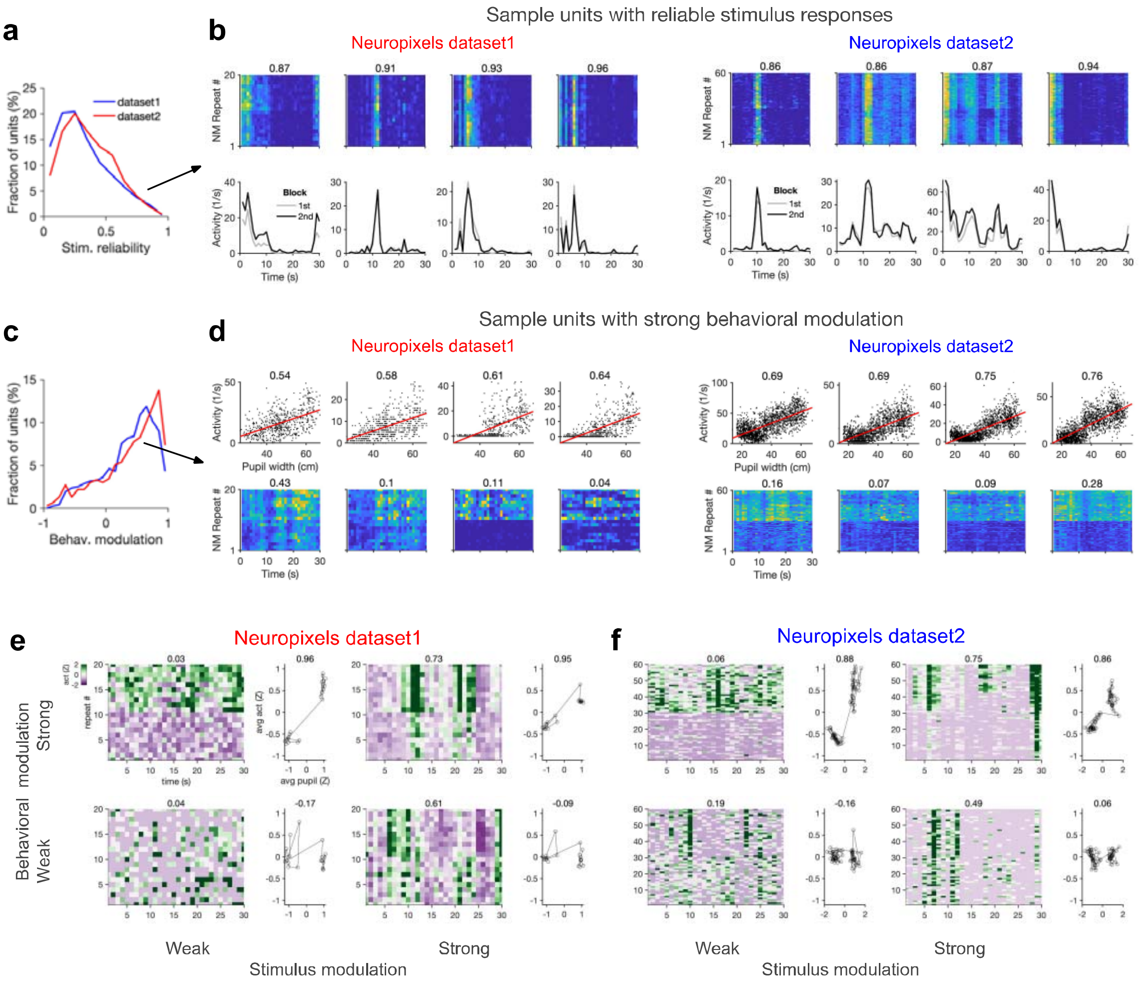
Extended Data Fig. 3: Dependence of representational similarity on behavioural change in wild type mice and for male/female animals separately.

Same as Extended Data Fig. 2 when (**a**) only wild typed (WT) mice are included in the analysis, or (**b**) when the analysis is performed for V1 units in female and male mice separately (see Supplementary Table 1 for details). Left: Neuropixels dataset1; Right: Neuropixels dataset2.

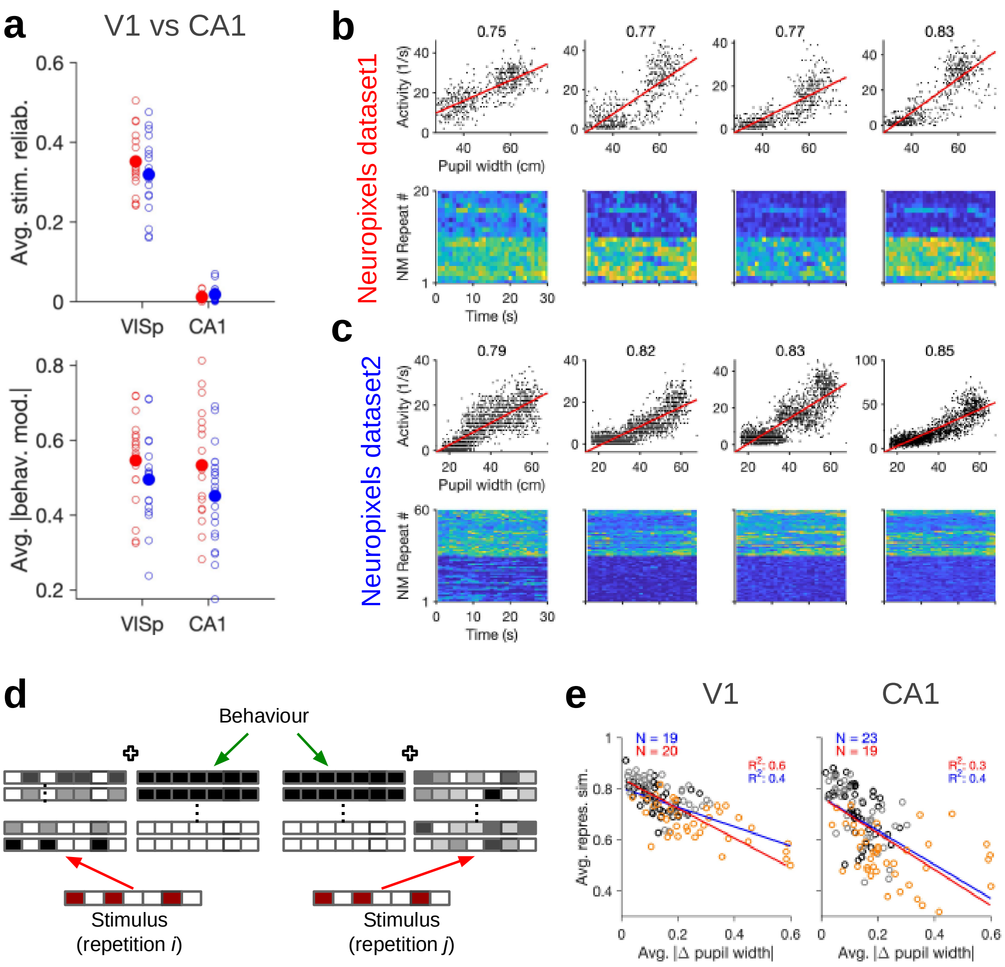


Extended Data Fig. 4: Dependence of representational similarity on behaviour in different gain models and in different experimental datasets.

a, A model neuron which integrates the signal and the noise components in its inputs. The signal has the same pattern over multiple repetition (rep#) of the stimulus, while the noise changes in each repeat. **b**, Representational similarity as a function of relative gain of the repeats (g_i/g_j, where g_i and g_j are the gains in the i-th and j-th repeats) for three models, where both signal and noise (left), only noise (middle), or only signal (right) components of the input are scaled by behaviour (see Methods for details). Change in the gain did not change representational similarity when both signal and noise were scaled (left), consistent with our theoretical analysis (see Methods). When arousal scaled noise only, there was a small decrease in the average representational similarity (middle). The most prominent effect was observed when arousal scaled signal only. For this scenario, a general increase in the average representational similarity was obtained, with the maximum increase happening at equal gains ($g_i = g_j$) (right). **c**, Representational similarity as a function of relative pupil size (obtained by the division of the average pupil sizes in a pair of movie repeats) for all recorded units. **d**, The average representational similarity of all mice shown in (**c**) for datasets 1 (red) and 2 (blue) separately. Bottom: Same when V1 units are only included, from the sessions with more than 40 units (the inclusion criterion). **e**, Same as (c) when V1 units are only included in the analysis. There are fewer individual sessions here because not all sessions contained more than 40 V1 units. **f**, Same as (d) for V1 units.

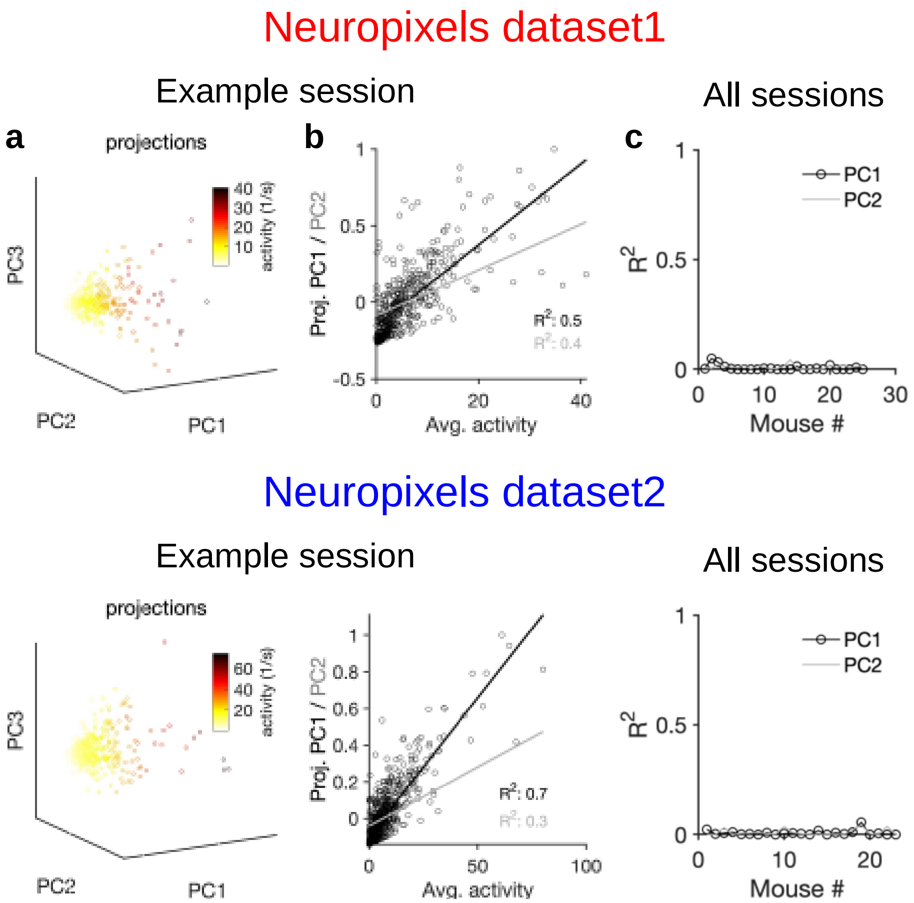


Extended Data Fig. 5: Wide and mixed distribution of stimulus and behavioural modulations. **a**, Distribution of stimulus reliability for all V1 units from all sessions in Neuropixels dataset1 (red) and dataset2 (blue). **b**, Sample activity of V1 units with high stimulus reliability (indicated by the numbers on the top) from each dataset. Top: The activity in response to each movie repeat; bottom: average activity in each block of presentation. **c**, Distribution of behavioural modulation of all V1 units for the two datasets. Behavioural modulation is obtained as the correlation coefficient (CC) of each unit's activity with pupil size. **d**, Sample activity of V1 units with strong modulation by pupil size (numbers indicated on the top). Top: tuning of unit's activity with pupil size. Bottom: the activity of units in response to repeats of the natural movie, showing different levels of modulation by stimulus within and across blocks of presentation (denoted by the value of stimulus reliability on top). **e,f**, Activity of units can be weakly or strongly modulated by stimulus or behaviour, giving rise to four possible quadrants. Sample V1 units from each quadrant are shown for Neuropixels dataset1 (e) and dataset2 (f). For each sample, z-score activity of the unit across different repetitions of the movie is plotted (left), with the number on top denoting stimulus reliability of the unit. The average activity of each unit as a function of average pupil size (during each movie repeat) is plotted on the right, with the number on the top denoting behavioural modulation of the unit (CC with pupil size).



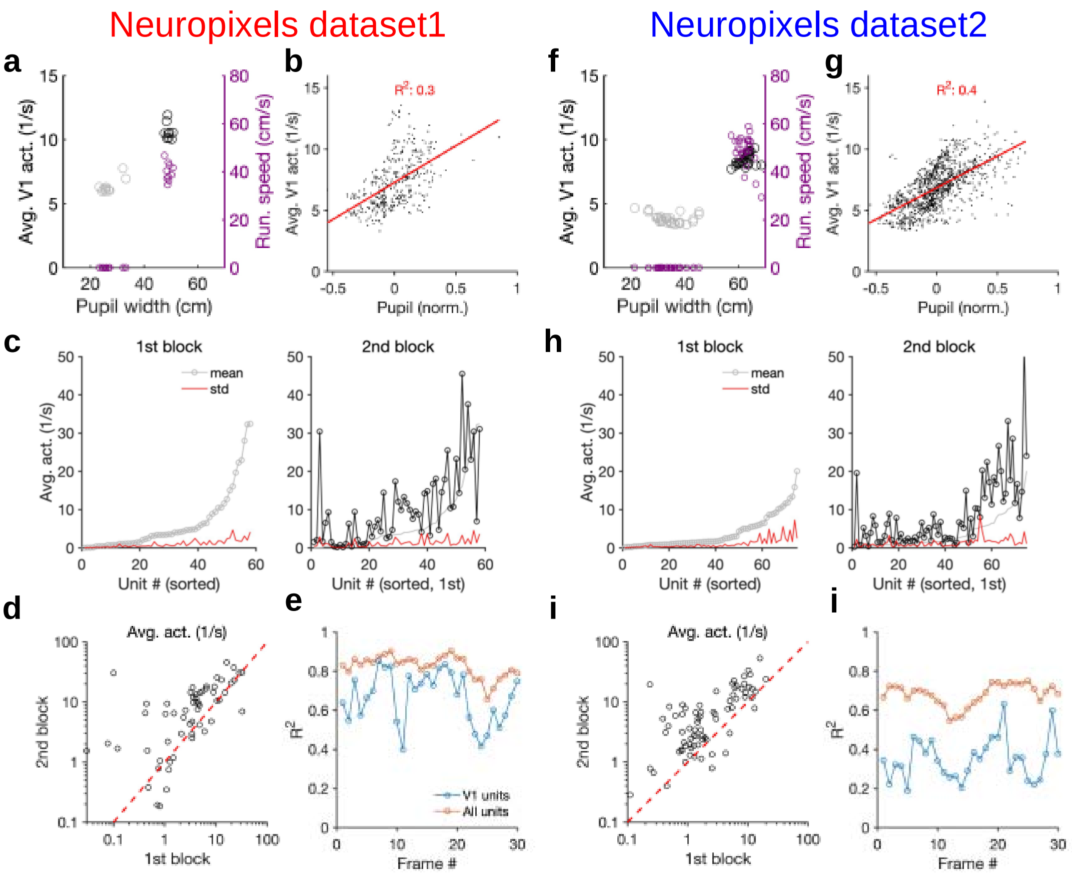
Extended Data Fig. 6: Stimulus-independent behavioural modulation of CA1 units.

a, Top: Average stimulus reliability across units in V1 and CA1 for different mice in each dataset. Bottom: Same for the average (across units) of the absolute value of behavioural modulation. Filled circles: average across mice. Red: Neuropixels dataset1; Blue: Neuropixels dataset2. **b**, Top: Sample activity of CA1 units (from Neuroixels dataset1) with considerable modulation by pupil size (numbers indicated on the top). Bottom: The activity of units in response to repeats of the natural movie. **c**, Same as (**b**) for Neuroixels dataset2. **d**, Schematic representation of population responses with stimulus-evoked (red) and behaviourally-induced (green) components to the repeats of the same stimulus. Even if the stimulus-evoked component is different between repeats (red), the population vector of responses (see Extended Data Fig. 1c) can have some similarity due to the constancy of the component set by the behaviour (green). **e**, Average representational similarity as a function of change in pupil width (similar to Fig. 1d,h, right) for V1 (left) and CA1 units (right). Red: Neuropixels dataset1; Blue: Neuropixels dataset2.



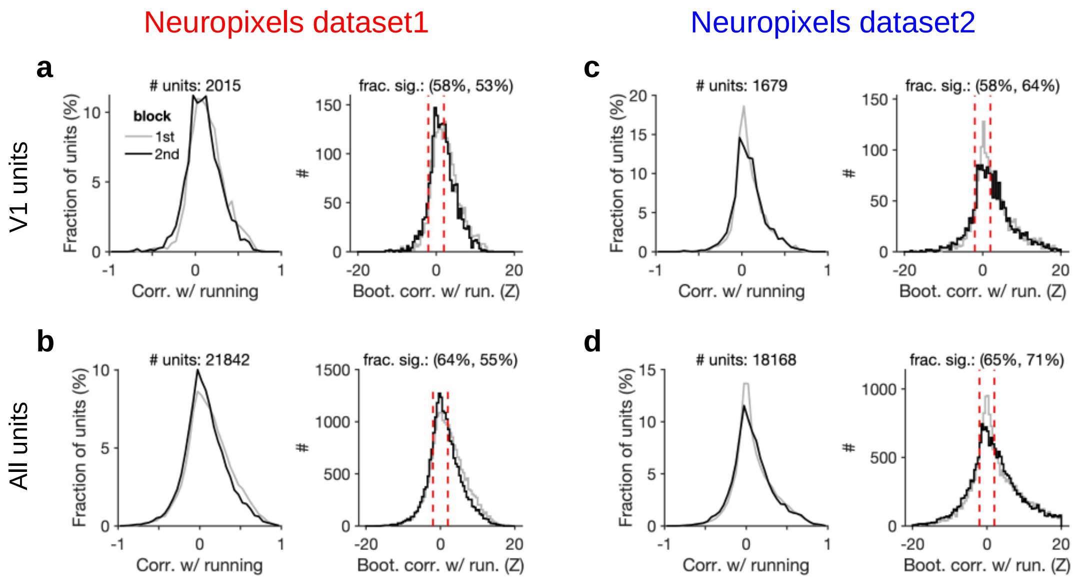
Extended Data Fig. 7: Relation of the principal components of neural activity to the average activity of units and their stimulus reliability.

a, Projections of the activity of units in example sessions over the first three PCs (cf. Fig. 3a,b,g,h), with the average activity of each unit indicated by the pseudo colour code. **b**, Projection of units' activity over PC1/PC2 versus the average activity of the unit. The best fitted regression lines and R^2 values in each case are shown. **c**, Similar to Fig. 3c,i for individual sessions. R^2 values of regression lines fitted to the projection of units' activity over PC1/PC2 versus stimulus reliability of the respective units is plotted for each session/mouse. Upper: Neuropixels dataset1; Lower: Neuropixels dataset2.



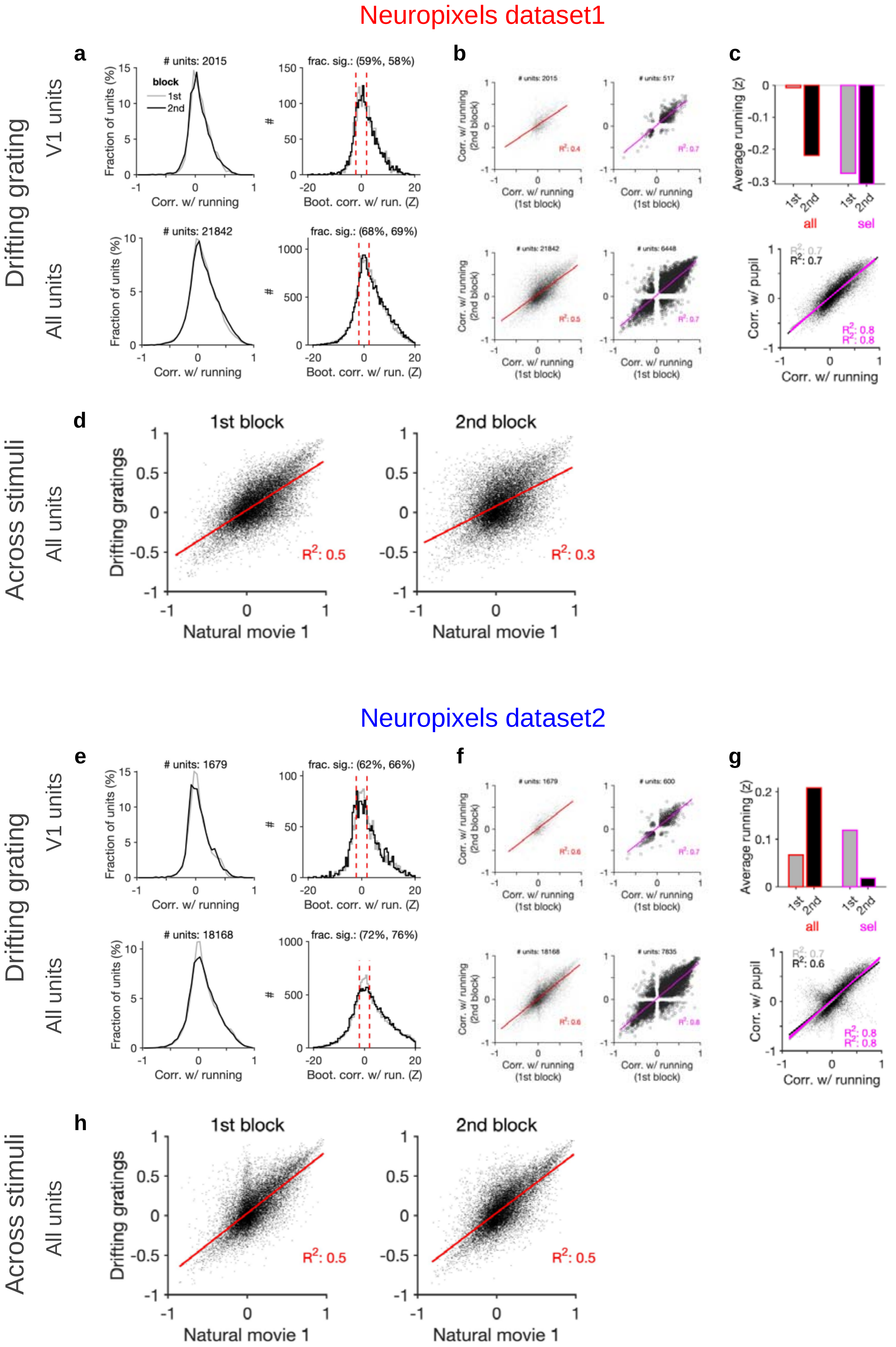
Extended Data Fig. 8: Average activity of units is modulated by behavioural state.

a, Average population activity (left y-axis) and running speed (right y-axis) as a function of pupil size, for the example shown in Fig. 4a from Neuropixels dataset1. **b**, Average population activity of V1 units during each movie presentation as a function of pupil size, from all recorded sessions. Pupil size for each repeat is normalized (within each session) by subtracting the mean value (across repeats) and dividing by it. **c**, For the example session in Fig. 4, the average (across movie frames) activity of V1 units is calculated and their mean and std across movie repetitions in each block is shown. Units are sorted in both blocks according to the mean in the 1st block. **d**, Average activity (across movie frames and repeats) of units during the 2nd block versus the 1st. Note the logarithmic scales. **e**, R^2 values of the regression fits to the data like Fig. 4c, when the population vectors are composed of the average activity of units during presentation of each individual frame (1 second long) of the natural movie. **f-j**, Same as (**a-e**) for Neuropixels dataset2.



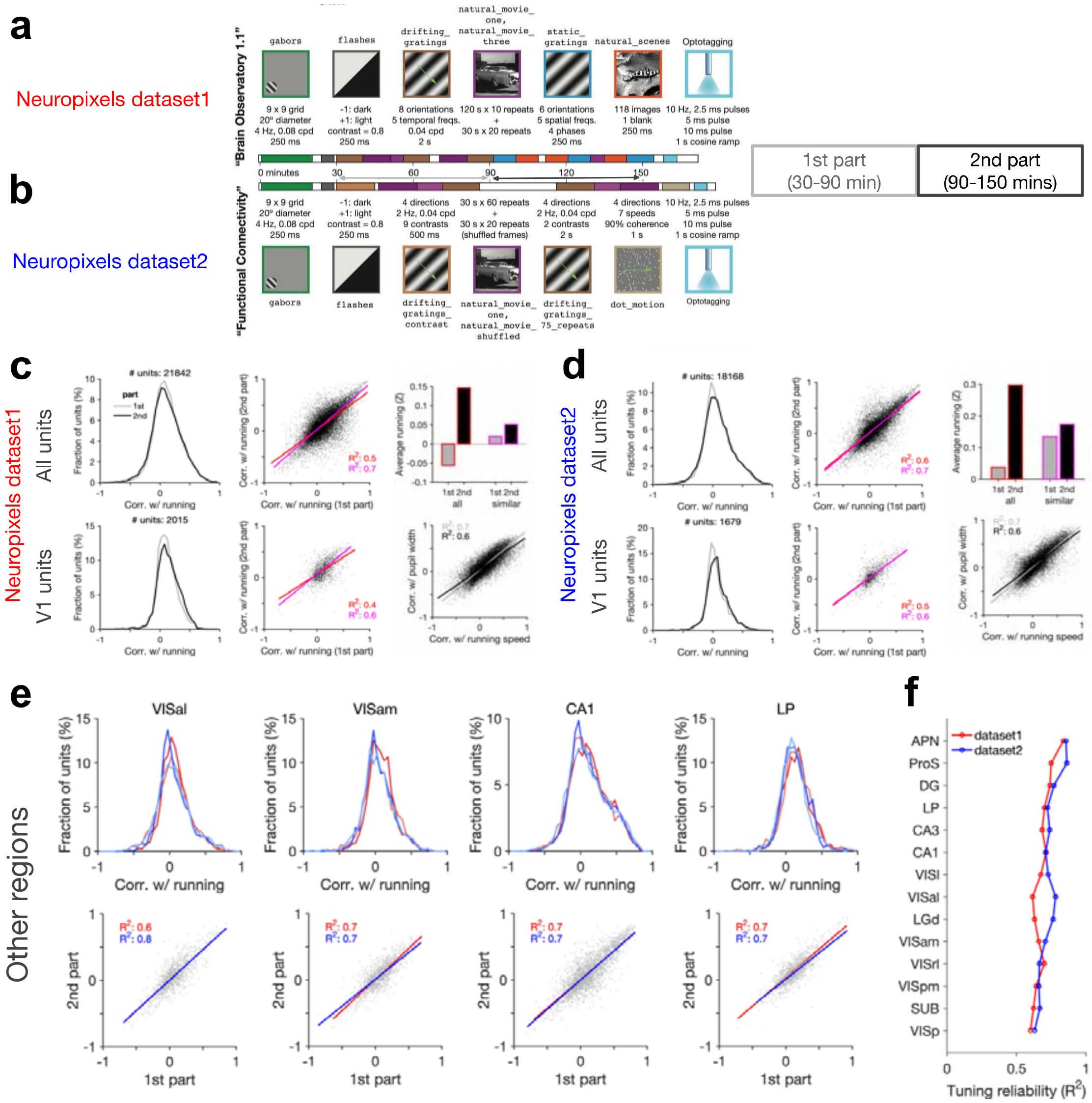
Extended Data Fig. 9: Significant modulation of units by behaviour.

a, Left: Distribution of correlation of V1 units' activity with running during first and second blocks of presentation of natural movie 1. Right: Distribution of bootstrapped correlations with running. Correlation coefficient (CC) of each unit's activity with 100 randomly shuffled versions of the running speed is calculated. The z-score of bootstrapped correlation (Z) is calculated by subtracting the mean of this distribution from the unshuffled CC and dividing it by the std of the distribution (see Methods for details). Bootstrapped correlations are calculated during the first (grey) and second (black) blocks separately. Significant correlations are taken as units for which |Z| > 2 (indicated by dashed red lines). Fractions of significant correlations during the first and second blocks are indicated on the top, respectively. **b**, same as (**a**) for all recorded units. **c,d**, Same as (**a,b**) for dataset2.


Neuropixels dataset1
a
b
c
Drifting grating
V1 units
All units
# units: 2015
block
1st
2nd
Fraction of units (%)
Corr. w/ running
frac. sig.: (59%, 58%)
#
Boot. corr. w/ run. (Z)
# units: 21842
frac. sig.: (68%, 69%)
# units: 517
# units: 6448
Corr. w/ running (2nd block)
Corr. w/ running (1st block)
R²: 0.4
R²: 0.7
R²: 0.5
R²: 0.7
Average running (z)
1st
2nd
all
sel
Corr. w/ pupil
R²: 0.7
R²: 0.7
R²: 0.8
R²: 0.8
d
Across stimuli
All units
1st block
2nd block
Drifting gratings
Natural movie 1
R²: 0.5
R²: 0.3
Neuropixels dataset2
e
f
g
# units: 1679
frac. sig.: (62%, 66%)
# units: 18168
frac. sig.: (72%, 76%)
# units: 600
# units: 7835
R²: 0.6
R²: 0.7
R²: 0.6
R²: 0.8
R²: 0.7
R²: 0.6
R²: 0.8
R²: 0.8
h
R²: 0.5
R²: 0.5

Extended Data Fig. 10: Consistent modulation of neuronal responses by behaviour across blocks of presentation of drifting gratings, and across stimuli.

a, Left: Distribution of correlations with running during the 1st and 2nd blocks of presentation of drifting gratings across all sessions. Right: Distribution of the z-score of bootstrapped correlations (Z) with running (see Methods and Extended Data Fig. 9). Significant correlations with running are defined as |Z| > 0.2. **b**, Correlation with running of units during the 2nd block against the 1st block, for all units and sessions (left; red), and for selected units (right; magenta), where sessions with similar levels of running between the two blocks and units with significant correlations are selected. **c**, Upper: Average running during the 1st and 2nd blocks for all sessions (all; red) and for selected units (sel; magenta). Lower: Correlation of all units with pupil versus their correlation with running, during the 1st (grey) and 2nd (black) blocks. Magenta: regression fits for selected units only. **d**, Correlation of all units with running speed during the presentation of drifting gratings versus correlations with running obtained during the presentation of natural movie 1, in the 1st and 2nd blocks of presentations, respectively. **e-h**, Same as (**a-d**) for dataset2.

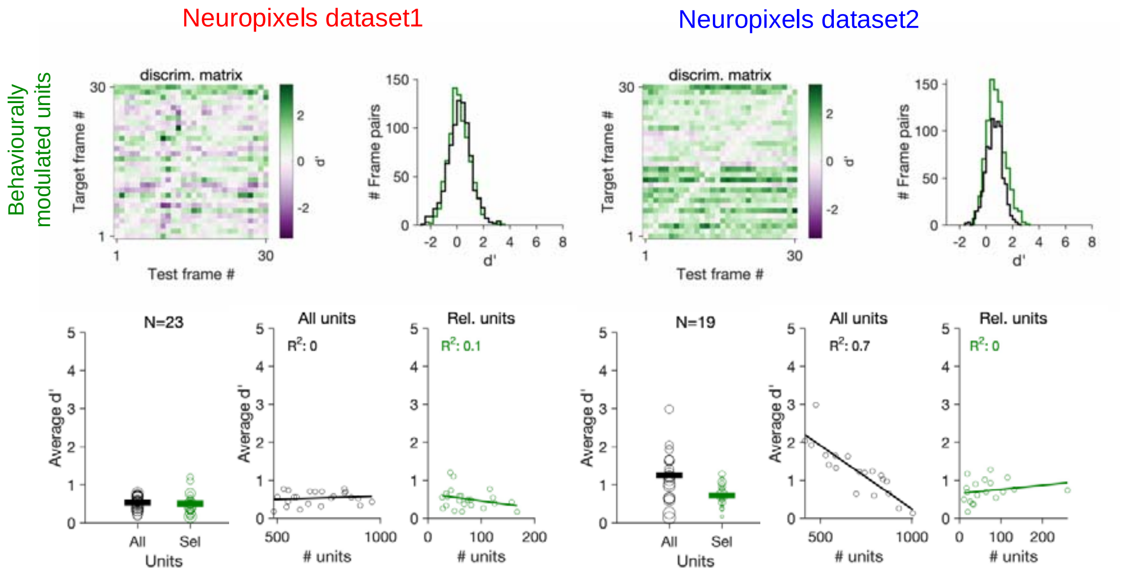


Extended Data Fig. 11: Consistent modulation of neuronal responses by behaviour across stimuli, regions and datasets.

a,b, Analysis of the reliability of behavioural tuning in two parts of each session in the two datasets. The composition of stimulus sets in each session type is shown, with the type, sequences, and the length of each stimulus presentation indicated[1]. In both datasets, correlation of units with running speed is calculated in two parts: 1st part from 30 to 90 minutes, and the 2nd part from 90 to 150 minutes. **c,d**, Similar to Extended Data Fig. 10a-c for the 1st and 2nd part of the sessions in dataset1 (**c**) and dataset2 (**d**). **e**, Upper:

[1] Illustrations from: https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html (for further details, see: https://allensdk.readthedocs.io and http://observatory.brain-map.org/visualcoding).

Distribution of correlations with running for units recorded from different regions. Results for the two parts of the sessions (lighter lines denoting the first part) and both datasets (red: dataset1; blue: dataset2) are overlayed. Sessions where the average running between the two parts are too different are excluded (exclusion criteria: |Z2-Z1|>0.3, where Z1 and Z2 are the average of the z-scored value of running speed in the 1st and 2nd parts, respectively). Lower: Correlation with running in the 2nd part against the 1st part in each region, for dataset1 (red) and dataset2 (blue), respectively. Lines show the best fitted least-square regression lines, with numbers denoting the R^2 values of the fit in each case. **f**, Tuning reliability (average R^2 values in (**e**)), for different regions across the two datasets. Regions key: [visual cortex, VIS] VISp: primary visual cortex; VISl: lateromedial area; VISrl: rostrolateral area; VISal: anterolateral area; VISam: posteromedial area; VISam: anteromedial area. [Hippocampal formation] CA1: cornu ammonis 1; CA3: cornu ammonis 3; DG: dentate gyrus; SUB: subiculum; ProS: prosubiculum. [Thalamus] LGd: lateral geniculate nucleus; LP: lateral posterior nucleus. [Midbrain] APN: anterior pretectal nucleus.



Extended Data Fig. 12: Decoding natural images does not improve by focusing on behaviourally modulated units.

Same as Fig. 7c,d when reliable (Rel.) units are chosen as units with strong behavioural modulation (correlation with running speed of more than 0.5), instead of units with strong stimulus reliability (cf. Fig. 7). Relation between average d’ and the number of units available for decoding in each session (all units or behaviourally reliable units) is plotted on the bottom.

Supplementary Table 1: Information of recording sessions in different datasets.

SESSION ID	SESSION TYPE	AGE (DAY)	SEX	GENOTYPE	PUPILOMETRY
715093703	brain observatory	118	M	Sst-IRES-Cre x Ai32	NO
719161530	brain observatory	122	M	Sst-IRES-Cre x Ai32	NO
721123822	brain observatory	125	M	Pvalb-IRES-Cre x Ai32	NO
732592105	brain observatory	100	M	WT	NO
737581020	brain observatory	108	M	WT	NO
739448407	brain observatory	112	M	WT	NO
742951821	brain observatory	120	M	WT	YES
743475441	brain observatory	121	M	WT	YES
744228101	brain observatory	122	M	WT	YES
746083955	brain observatory	98	F	Pvalb-IRES-Cre x Ai32	YES
750332458	brain observatory	91	M	WT	YES
750749662	brain observatory	92	M	WT	YES
751348571	brain observatory	93	F	Vip-IRES-Cre x Ai32	YES
754312389	brain observatory	140	M	WT	YES
754829445	brain observatory	141	M	WT	YES
755434585	brain observatory	100	M	Vip-IRES-Cre x Ai32	YES
756029989	brain observatory	96	M	Sst-IRES-Cre x Ai32	YES
757216464	brain observatory	105	M	WT	YES
757970808	brain observatory	106	M	WT	YES
758798717	brain observatory	102	M	Sst-IRES-Cre x Ai32	YES
759883607	brain observatory	113	M	WT	YES
760345702	brain observatory	103	M	Pvalb-IRES-Cre x Ai32	YES
760693773	brain observatory	110	F	Sst-IRES-Cre x Ai32	YES
761418226	brain observatory	119	M	WT	YES
762120172	brain observatory	100	M	Vip-IRES-Cre x Ai32	YES
762602078	brain observatory	110	M	Sst-IRES-Cre x Ai32	YES
763673393	brain observatory	126	M	WT	YES
773418906	brain observatory	124	F	Pvalb-IRES-Cre x Ai32	YES
791319847	brain observatory	116	M	Vip-IRES-Cre x Ai32	YES
797828357	brain observatory	107	M	Pvalb-IRES-Cre x Ai32	YES
798911424	brain observatory	110	F	Vip-IRES-Cre x Ai32	YES

799864342	brain observatory	129	M	WT	YES
766640955	functional connectivity	133	M	WT	YES
767871931	functional connectivity	135	M	WT	YES
768515987	functional connectivity	136	M	WT	NO
771160300	functional connectivity	142	M	WT	YES
771990200	functional connectivity	108	M	WT	YES
774875821	functional connectivity	114	M	WT	YES
778240327	functional connectivity	120	M	WT	YES
778998620	functional connectivity	121	M	WT	YES
779839471	functional connectivity	122	M	WT	YES
781842082	functional connectivity	126	M	WT	YES
786091066	functional connectivity	111	F	Sst-IRES-Cre x Ai32	YES
787025148	functional connectivity	114	M	Sst-IRES-Cre x Ai32	YES
789848216	functional connectivity	119	M	Sst-IRES-Cre x Ai32	YES
793224716	functional connectivity	120	M	WT	YES
794812542	functional connectivity	120	F	Sst-IRES-Cre x Ai32	YES
816200189	functional connectivity	128	F	Vip-IRES-Cre x Ai32	YES
819186360	functional connectivity	128	F	WT	YES
819701982	functional connectivity	135	F	Vip-IRES-Cre x Ai32	YES
821695405	functional connectivity	134	F	WT	YES
829720705	functional connectivity	112	M	Pvalb-IRES-Cre x Ai32	YES
831882777	functional connectivity	137	M	Sst-IRES-Cre x Ai32	YES
835479236	functional connectivity	121	M	Vip-IRES-Cre x Ai32	YES
839068429	functional connectivity	129	F	Sst-IRES-Cre x Ai32	YES
839557629	functional connectivity	115	M	Pvalb-IRES-Cre x Ai32	YES
840012044	functional connectivity	116	M	Pvalb-IRES-Cre x Ai32	NO
847657808	functional connectivity	126	F	WT	YES